Etudes par RMN des L,D-transpeptidases bactériennes : structure, dynamique et compréhension de leur inhibition par les beta-lactames

Lauriane Lecoq

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale**

Arrêté ministérial : 7 août 2006

Présentée par

Lauriane Lecoq

Thèse dirigée par **Jean-Pierre Simorre**
et codirigée par **Bernhard Brutscher**

préparée au sein de **Biomolecular NMR group, Institut de Biologie Structurale**
et de **Ecole Doctorale Chimie et Science du Vivant**

Etude par RMN des L,D-transpeptidases bactériennes: structure, dynamique et caractérisation de leur inhibition par les β-lactames

Thèse soutenue publiquement le **29 Novembre 2012**,
devant le jury composé de :

Pr. Winfried Weissenhorn
Professeur à l'Université Joseph Fourier, Grenoble, Président

Pr. Bruno Kieffer
Professeur à l'Université de Strasbourg, Rapporteur

Dr. Ivo Gomperts Boneca
Docteur à l'Institut Pasteur, Paris, Rapporteur

Dr. Michel Arthur
Docteur au Centre de Recherche des Cordeliers, Paris, Examinateur

Dr. Jean-Pierre Simorre
Docteur à l'Institut de Biologie Structurale, Grenoble, Directeur de thèse

Dr. Catherine Bougault
Maître de conférences à l'Université Joseph Fourier, Grenoble, Co-Directeur de thèse

Remerciements

Je tiens tout d'abord à remercier mes deux encadrants, Jean-Pierre et Cathy, pour tout ce qu'ils m'ont appris durant ces trois années. Cathy, merci pour ton soutien sans faille et pour l'immense générosité dont tu fais preuve pour accompagner tous tes étudiants, dont j'ai eu la chance de faire partie. Un grand merci également pour les nombreuses heures passées à corriger mon manuscrit! Jean-Pierre je te remercie pour m'avoir donné l'opportunité de prendre part à cette aventure, et pour me donner envie de toujours aller plus loin. Les discussions qui nous ont animés tous les 3 auront toujours été très productives, et j'espère pouvoir mettre à profit dans le futur tout ce que vous m'avez apporté, tant sur le plan humain que sur le plan scientifique.

Je tiens tout particulièrement à remercier également les rapporteurs Bruno Kieffer et Ivo Boneca, qui ont accepté d'évaluer mon travail. En espérant qu'ils le trouvent intéressant et qu'il en résulte de fructueuses discussions.

Je souhaite remercier tous les membres de l'équipe, sans qui les idées scientifiques ne fuseraient sans doute pas autant, et surtout pour m'avoir donné envie d'aller chaque jour au travail avec le sourire. Merci à Bernhard pour m'avoir donné l'opportunité de faire partie de cette équipe, et pour toujours rester si simple malgré les responsabilités du groupe. Merci à Adrien et à Beate pour leur précieuse aide lors des laborieux calculs de structure, et pour avoir toujours été disponibles pour répondre à mes questions. Adrien, que ferait le groupe sans ton aide pour les spectros et pour régler les problèmes informatiques?! Je n'oublierai pas les nombreuses heures que tu m'as épargnées. Pierre, même si tu sais bien que je voudrais t'interdire de venir au labo avec tes habituelles chaussettes dans tes sandales, j'ai beaucoup apprécié nos discussions et nos taquineries. J'espère que Nicolas saura prendre le relai pour t'embêter une fois que je serai partie. Dominique, un grand merci à toi pour ton partage de connaissances de RMN, tout comme ton partage sur tes expériences Guadeloupéennes qui nous ont fait rêver plus d'une fois (ahh les habituelles photos de poisson exotiques à la fin des présentations powerpoint...). Enfin, le groupe ne serait pas complet sans citer Jérôme et ses idées scientifiques toujours très affutées et Eric pour ses t-shirt mythiques. Ce fut un réel plaisir de travailler et d'interagir avec tous les membres du groupe, qui j'espère restera toujours aussi soudé et enrichissant.

Malgré leur isolement au CIBB, on ne peut pas non plus oublier Isabel, Ombeline et Carine. Lorsque je suis arrivée à l'IBS, l'on m'avait dit qu'Isabel avait des pouvoirs magiques pour la préparation des échantillons biologiques. Je n'ai pas été déçue, puisque la belle LdtMt1 nous est parvenue soluble et concentrée! Merci pour ton enthousiasme et ton indéniable talent. Merci à Ombeline et à Carine pour avoir participé à la production des échantillons, et nous permettre ainsi de passer plus de temps devant le spectro.

Lors de ces trois années, j'ai également eu la chance de faire partie d'un groupe de travail grâce au projet ANR et de pouvoir interagir avec Michel Arthur et son groupe au Centre de Recherche des Cordeliers à Paris. Je garde un très bon souvenir des différentes réunions ANR, qui nous ont réunis tour à tour à Méaudre, à Paris, et à Bourbon, car elles m'ont permis à chaque fois de participer à des échanges

scientifiques particulièrement enrichissants, et de mieux comprendre la portée de mon travail de thèse au sein d'un projet de plus grande ampleur. Un grand merci donc à Michel, Carole, Manu, Vincent et Sébastien, pour avoir permis ces échanges, tout ceci dans une ambiance toujours chaleureuse. Dans ce contexte, je remercie également Marie-Etheve Quelquejeu et Hélène pour la synthèse du nouveau carbapenem beta32, et pour nous avoir fait (re)découvrir les notions de chimie de synthèse au cours de ces réunions.

En parallèle de mon projet sur les LDts, j'ai eu l'opportunité de travailler en collaboration avec Vincent, Christel et Babeth du Laboratoire de Chimie et Biologie des Métaux, que je tiens tout particulièrement à remercier pour m'avoir fait découvrir le monde des cellules, et pour les bons moments passés. Lors de ces expériences, j'ai été en contact avec Daphna pour les mesures de microscopie électronique, que je remercie vivement pour sa gentillesse.

Je remercie l'organisme qui finance ma thèse, l'Agence Nationale de la Recherche (ANR), ainsi que les organismes qui ont participé au financement de différents voyages consacrés à des collections de données RMN sur des plateformes nationales (à Gif-sur-Yvette, avec le TGIR) et internationales (à Florence, avec EUNMR). Sans ces organismes, le partage scientifique ne pourrait être aussi développé.

Enfin, je remercie évidemment tous les thésards, post-docs et stagiaires, qui font que le temps passé au labo soit toujours agréable, et que le temps passé hors labo le soit tout autant. Je pense notamment à Zsofia, Mathieu P., Nicolas, Guillaume, Mathilde et Jens pour les sorties ski / via ferrata / laser game et autres. J'espère retrouver des gens motivés comme vous dans mon futur labo! Je pense également à Paul, Mélanie, Michael, Pavel, Rime, Matthieu et Enrico, ainsi que tous ceux qui sont déjà partis mais pour qui j'ai une pensée aujourd'hui: Maria pour ses bons gâteaux lors des réunions spectros, et Sophie pour son rire inoubliable qui s'entendait d'un bout à l'autre du labo. Je n'oublie pas non plus les membres du FDP, toujours là pour aller boire un coup: Valéry, Loïc, Luca, Robert, Malène et Guillaume.

Je vais clôturer cette liste non exhaustive en remerciant ma famille et mes amis. Mes parents tout d'abord, qui ont toujours cru en moi, m'ont encouragée et m'ont toujours poussée à faire ce dont j'avais le plus envie. Je ne serais sans doute pas allée aussi loin sans votre confiance. Ma soeur Laetitia et mon frère Yannick car je sais que je pourrai toujours compter sur vous. Je souhaite remercier également mes amis proches, notamment mes amis de lycée avec qui j'ai la chance d'avoir gardé des liens très forts: Aliénor, Jérémie, Hugo, Fabien, Elodie, Lilice et Astrid; c'est toujours un immense plaisir de vous retrouver. Egalement mes amis faluchards, auprès de qui j'ai toujours pu trouver confiance, un peu de folie et beaucoup de complicité, et avec qui j'ai partagé des moments inoubliables: Alice, Léo, Oanellig, Quentin, Emi, Nico, Floran, Hélène, Coline, Mathilde, Anaïd et Juliette (et pardon si j'en oublie...). Vous allez tous extrêmement me manquer! Un grand merci enfin, à Ariane, pour être à mes côtés chaque jour depuis 4 ans et avec qui l'aventure va maintenant pouvoir se poursuivre au Canada :-)

General introduction

β-lactams are the most widely used drugs for the treatment of bacterial infections. This antibiotic class comprises among others penicillins, cephalosporins, and carbapenems, which have been successively developed to fight antibiotic resistance. Until recently, all β-lactams were thought to exclusively act on the active-site serine D,D-transpeptidases (PBPs) that catalyze the final cross-linking step of the peptidoglycan synthesis, peptidoglycan being the main component of the bacterial cell wall. It was shown recently that these D,D-transpeptidases can be by-passed by a novel class of enzymes, the L,D-transpeptidases (LDts), in β-lactam-resistant enterococci and in *Mycobacterium tuberculosis*. These enzymes are promising targets for the development of anti-tuberculous drugs belonging to the carbapenem class since this antibiotic family inactivates the L,D-transpeptidases that are predominantly responsible for the *M. tuberculosis* peptidoglycan cross-linking.

The present work focuses on this new class of enzymes and the mechanism of their inhibition by the carbapenem β-lactam family.

The introduction (Part I) gives an overview of the biological context, with a general presentation of the bacterial cells composition and with a description of the existing antibiotics and their corresponding protein targets within the cell (Chapter 1). The second chapter included in the introduction subsequently describes in more details the peptidoglycan component of bacterial cell wall, with a description of its physicochemical characteristics and of its biosynthesis within the cytoplasm and the periplasm compartments of the cell. A particular emphasis is given to the transpeptidation reaction that cross-links the peptide stems covalently linked to the glycan chains of the growing peptidoglycan. This reaction can be alternatively catalyzed by D,D-transpeptidases (DDts) or L,D-transpeptidases (LDts).

The second part presents the NMR methods used for the characterization of the interaction between LDts and β-lactam antibiotics, and is divided into three chapters. First the structural aspects of the studies are detailed in Chapter 3. NMR methods used from the resonance assignment to the structure determination of an enzyme and an enzyme/inhibitor complex are described. Approaches developed to study high- versus low-affinity complexes are distinguished. Secondly, the extraction of thermodynamical parameters from NMR data is presented in Chapter 4. It includes procedures to determine the dissociation constants K_D that characterize the strength of the protein-ligand interaction, and procedures to extract the pKa values of ionizable residues in proteins. A particular emphasis is given to the case of histidines, as they are involved in a large number of catalytic processes, including the LDt acylation reaction, and as their imidazole ring can exist as three different forms in solution, instead of two for the other ionizable residues. Thirdly, NMR can provide insights into the dynamics of enzymes and of enzymes/inhibitor complexes, as described in Chapter 5. Measurements of relaxation data are exposed, together with the extraction of relaxation rates and their analysis to characterize fast and slower motion time-scales as well as conformational exchange contributions.

The third and last part of this work includes the results obtained on the LDts from three different bacterial strains. Chapter 6 summarizes the structure and dynamics of the LDt from *Bacillus subtilis* ($\mathrm{Ldt_{Bs}}$) within a publication in the journal Structure. Chapter 7 describes the behavior of the LDt from *Enterococcus faecium* ($\mathrm{Ldt_{fm}}$) as well as preliminary studies on one of the five LDts from *Mycobacterium tuberculosis* ($\mathrm{Ldt_{Mt1}}$). A comparison between $\mathrm{Ldt_{Bs}}$ and $\mathrm{Ldt_{fm}}$ acylation mechanisms is provided in this chapter.

Finally, a general conclusion on this work will summarize the obtained results and our current understanding of the LDt acylation mechanism by carbapenems, as well as open new perspectives of research for a better apprehension of the essential activity of L,D-transpeptidases in *Mycobacterium tuberculosis*.

Contents

List of Abbreviations

AFM	Atomic Force Microscopy
AIR	Ambiguous Interaction Restraints
ARIA	Ambiguous Restraints for Iterative Assignment
ATP	Adenosine TriPhosphate
BMRB	Biological Magnetic Resonance Data Bank
CNS	Crystallography and NMR System
CPMG	Carr Purcell Meiboom Gill
CSA	Chemical Shift Anisotropy
CSP	Chemical Shift Perturbation
CT	Constant Time
DHPC	DiHexanoyl PhosphatidylCholine
DHPS	DiHydroPteroate Synthetase
DMPC	DiMyristoyl PhosphatidylCholine
TNB	2-nitro-5-thiobenzoate
DNA	DeoxyriboNucleic Acid
GCL	Glycan Chain Length
GlcNAc	*N*-acetyl-glucosamine
GT	GlycosylTransferase
HADAMAC	HADamard-encoded AMino-ACid
HADDOCK	High Ambiguity Driven protein-protein Docking
HMQC	Heteronuclear Multiple-Quantum Coherence
HPLC	High-Performance Liquid Chromatography
HSQC	Heteronuclear Single-Quantum Coherence
HMM	High Molecular Mass
IPTG	IsoPropyl-β-D-1-ThioGalactopyranoside
ITC	Isothermal Calorimetry
LDt	L,D-transpeptidase
Ldt$_{Bs}$	L,D-transpeptidase from *B. subtilis*
Ldt$_{fs}$	L,D-transpeptidase from *E. faecalis*
Ldt$_{fm}$	L,D-transpeptidase from *E. faecium*
Ldt$_{Mt1}$	L,D-transpeptidase 1 from *M. tuberculosis*
Ldt$_{Mt2}$	L,D-transpeptidase 2 from *M. tuberculosis*

LMM Low Molecular Mass

LTA LipoTeichoic Acid

MBLs Metallo-β-Lactamases

MurNAc *N*-Acetyl-Muramic Acid

MD Molecular Dynamics

MDR Multi-Drug Resistant

MIC Minimum Inhibitory Concentration

MRSA Methicillin-Resistant *Staphylococcus aureus*

NADPH DihydroNicotinamide-Adenine Dinucleotide Phosphate

NMR Nuclear Magnetic Resonance

NOE Nuclear Overhauser Effect

NOESY Nuclear Overhauser Effect SpectroscopY

OD Optical Density

OMP Outer Membrane Protein

PBP Penicillin-Binding Protein

PEG Poly(Ethylene Glycol)

PEP Phospho*enol*pyruvate

PDB Protein Data Bank

PNGM Phosphoglucosamine Mutase

RDC Residual Dipolar Coupling

RMSD Root Mean Square Deviation

RNA RiboNucleic Acid

SDS-PAGE Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis

SUMO Small Ubiquitin-related MOdifier

TA Teichoic Acid

TEM Transmission Electron Microscopy

TM TransMembrane

TP TransPeptidase

TOCSY TOtal Correlation SpectroscopY

TROSY Transverse Relaxation Optimized SpectroscopY

UDP Uridine DiPhosphate

WTA Wall Teichoic Acid

List of Figures

List of Tables

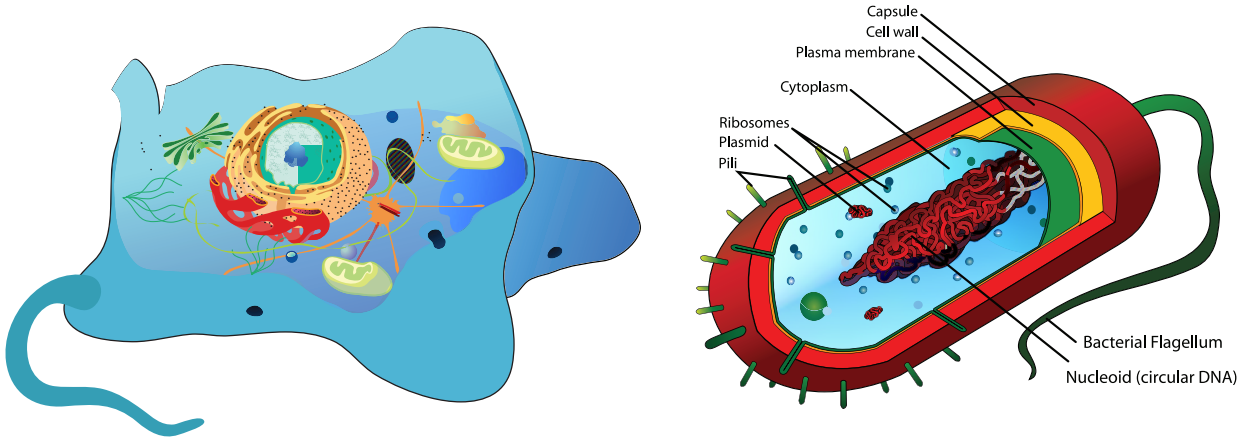
Part I

Introduction: Towards a new mechanism of bacterial antibiotic resistance

Chapter 1

Bacterial cells and existing antibiotics

A bacterium is a unicellular and prokaryotic microorganism. Conversely to eukaryotic organisms, characterized by the presence of a membrane-enclosed nucleus encapsulating their DNA and of diverse organelles, prokaryotes are devoid of a nucleus and in most cases of organelles in the cytoplasm, except ribosomes (figure 1.1). Eukaryotic cells (diameter 10 to 100 µm) generally have a thousand to a million times the volume of typical prokaryotes. Bacteria can nevertheless display a large range of size depending on the species, from about 0.1 µm for the mycoplasmas to 600 µm for *Epulopiscium fishelsoni*, which is genetically related to *Clostridium*. Typically, the averaged sized of *Escherichia coli* (*E. coli*) bacterial cell is about 2 µm long and 0.5 µm large. In addition, bacteria have a wide range of shapes, ranging from spheres (cocci) to rods (bacilli) and helical coils (spirilla). To add onto the bacterial diversity, the cell surface is in some cases covered with filamentous projection named pili, used for DNA transfer between cells or for attachment to host organism's cells. One or more appendages, known as flagella and used for locomotion, may also decorate the cell wall.



chapter in conjunction with involved cellular targets.

1.1 The bacterial cell anatomy

The bacterial cell can be roughly divided in two spaces, the cytoplasmic space and the cell envelope.

1.1.1 The cytoplasm

The bacterial cytoplasm contains a single chromosome (a DNA molecule warrant of the genetic information), several copies of which may be present at different growth stages. This chromosome, with a typical molecular weight of 2.5×10^9 kDa in *E. coli*, is condensed to form the nucleoid and represents typically 1% of the total *E. coli* molecular weight. Numerous RiboNucleic Acid (RNA) molecules and proteins respectively contribute to 6 and 15% of *E. coli* molecular weight. Part of the RNA and protein molecules are associated to form 250 Å-diameter ribonucleoprotein complex particles known as ribosomes, which are the centers of protein synthesis. Each prokaryotic ribosome is composed of a large 50S subunit containing 23S rRNA (2300 nucleotides), 5S rRNA (120 nucleotides) and 34 proteins, and of a small 30S subunit containing 16S rRNA (1540 nucleotides) and 21 proteins. Water, small organic molecules and inorganic ions complete the composition of the cytoplasm. Part of the cytoplasmic DNA occurs as plasmids that are not essential for the bacterial growth and can be transferred to other bacterial species as a result of horizontal gene transfer (Voet and Voet, 2004). Through recent advances in optical and electronic microscopy, a number of additional cytoplasmic structures or proteinaceous microcompartments that function like organelles in eukaryotes by sequestering specific enzymes have been evidenced (Murat *et al.*, 2010).

Such structures include:

- Filaments of cytoskeleton molecules that play an essential role in cell division (FtsZ), in cell elongation and polarity (MreB), in cell shape (Crescentin), in septation (MinCDE) (Vollmer, 2006)
- Nutrient storage structures such as sulfur granules, nitrate vacuoles or polyhydroxoalkanoates
- Carboxysomes, polyhedral bodies containing ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) associated within carbon dioxide fixation (Cheng *et al.*, 2008; Yeates *et al.*, 2008)
- Magnetosomes, membranous organelles that contain magnetite (Fe_3O_4) or greigite (Fe_3S_4) (Komeili, 2012)

1.1.2 The bacterial cell envelope

Located at the barrier between the cell and its environment, the bacterial cell envelope is essential for bacteria and their survival. It provides the cells with structural support and protection against external environmental attacks, and it acts as a filtering mechanism. The cell wall is thus considered as one of the major targets of antibiotics. One of its main components is the peptidoglycan, formerly known as murein. Through its contributions to counteract the internal osmotic pressure and to maintain the overall bacterial shape, peptidoglycan is the warrant of the cell stiffness and helps to maintain the overall shape of the bacterial cell.

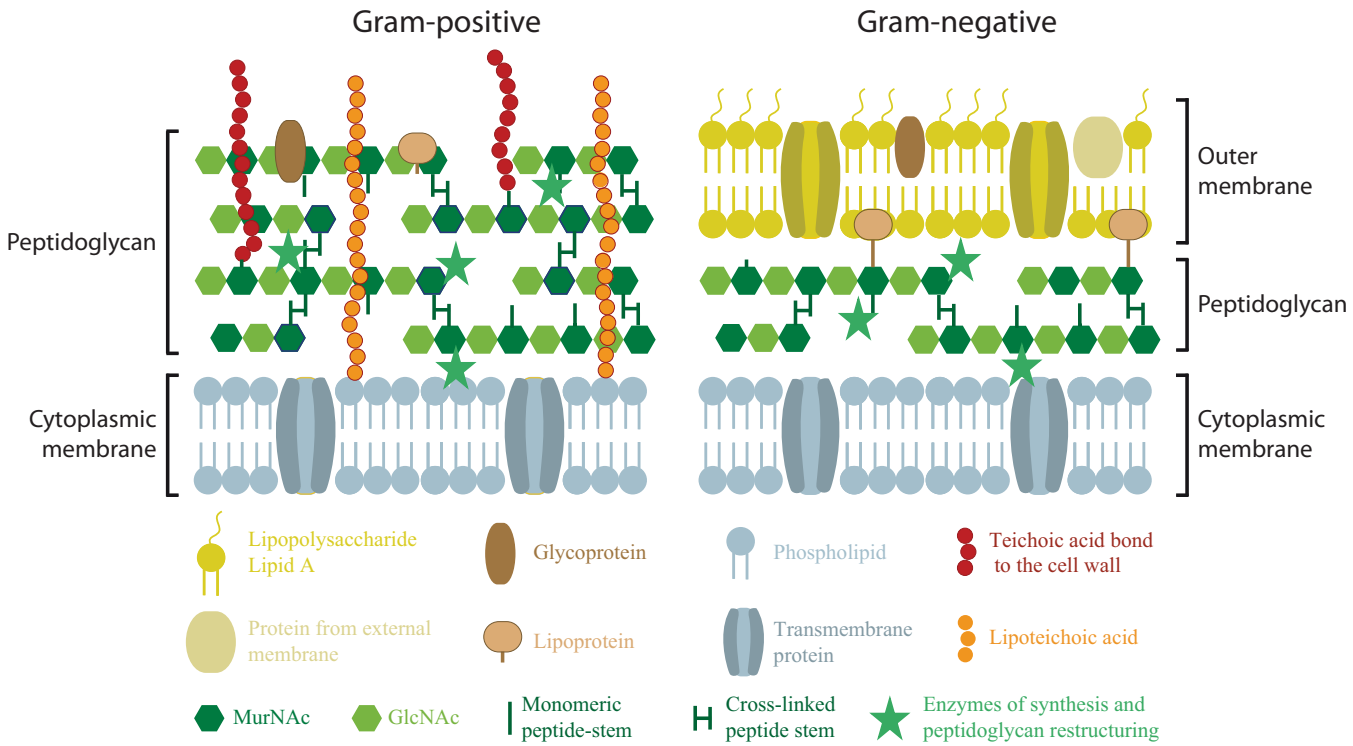


Figure 1.2: Schematic representation of the cell wall of Gram-positive and Gram-negative bacteria. Gram-positive bacteria have a thick layer of peptidoglycan containing teichoic acids. In contrast, the peptidoglycan in Gram-negative bacteria is very thin and protected by an outer membrane.

Gram coloration of cell wall and bacterial classification

Bacteria can be classified into two main categories, depending on their cell wall composition. "Gram-negative" bacteria have an envelope composed of an inner membrane (also called cytoplasmic membrane), a periplasmic space containing a single peptidoglycan layer, and an outer membrane that protects the peptidoglycan (figure 1.2). The following genera belong to this bacterial class: *E. coli*, the most studied bacterium to date, *Enterobacter*, *Pseudomonas* and *Haemophilus*. The cell wall of "Gram-positive" bacteria differs from the former class in the absence of outer membrane and in a thicker peptidoglycan, constituted by several layers (Matias and Beveridge, 2005). In the absence of an outer membrane, Gram-positive bacteria have developed alternative protective surface structures, such as capsular polysaccharides, S-layer proteins or mycolic acids in Actinobacteria. These features are highly variable and may be completely absent of the cell surface. One constant motif of Gram-positive cell envelopes, however, is the presence of additional glycopolymers, which are attached either to the peptidoglycan or to membrane lipids (see (Weidenmaier and Peschel, 2008) for a review). Among those cell-wall glycopolymers, teichoic acids are more certainly the best studied (see paragraph below). The following genera can be found in this Gram-positive bacterial class: *Bacillus*, among which the well established model species *Bacillus subtilis* (*B. subtilis*) in particular, *Staphylococcus*, *Enterococcus*, *Streptococcus*, *Clostridium*, *Listeria* and *Corynebacterium*.

The Gram stain, developed in 1884 by Hans Christian Gram, differentiates the two bacterial classes upon the structural features of their cell walls. The thick layers of peptidoglycan in the Gram-positive cell wall stain purple, while the thin Gram-negative cell wall appears pink. Yet, until 1951, almost nothing was known about the chemical composition of the bacterial cell wall. It was only after the link between β-lactams and cell wall was established that the interest of scientists for the latter (and peptidoglycan in particular), was revived. A few years later, a new method to isolate the cell wall was described by Salton *et al.* (Salton and Horne, 1951), opening the way to study this challenging system at a larger scale. Qualitative examinations of purified cell walls permitted to successfully elucidate the chemical structure of bacterial cell wall, especially into Gram-positive organisms.

Although they do not retain the violet coloration of the Gram stain due to a thicker peptidoglycan layer, the Actinobacteria phylum is classified as acid-fast Gram-positive bacteria as they do not possess an outer membrane. Among this phylum, the Mycobacterium genus includes pathogens known to cause serious diseases in mammals, including tuberculosis (*Mycobacterium tuberculosis*) and leprosy (*Mycobacterium leprae*). The peptidoglycan of mycobacterium has the particularity of being very thick, that is believed to result from a higher degree of cross-linking (Daffé and Reyrat, 2008). In fact, in *Mycobacterium tuberculosis* (*M. tuberculosis*), the cell wall accounts for up to 40% of the dry mass of cell, and the degree of cross-link is around 70 to 80% in *Mycobacterium* species (Gupta *et al.*, 2010), against ∼50% in the Gram-negative *E. coli* (Vollmer and Höltje, 2004).

The Gram-negative outer-membrane

The asymmetrical outer membrane of Gram-negative bacteria consists of phospholipids, lipopolysaccharides (composed of a hydrophobic membrane anchor lipidA, an oligosaccharide core region, and in some cases a repeating oligosaccharide), integral membrane proteins (with β-barrel structures in most cases) and lipoproteins (usually anchored to the membrane via an N-terminal N-acyl-diacylglycerylcysteine) (Bos *et al.*, 2007). These components are synthesized as precursors in the cytoplasm and processed into mature forms on the cytoplasmic membrane. They have then to be transported across the inner membrane and through the hydrophilic periplasm. Two translocation machineries across the inner membrane have been identified, namely the Sec system for unfolded proteins and the Tat system for folded proteins, while ABC transporters are used by the lipid components. Each component is then transported across the periplasm by dedicate chaperone or transporter proteins before to be assembled at the outer membrane (Okuda and Tokuda, 2011). The outer membrane functions as a selective barrier that protects the bacteria from harmful compounds such as antibiotics.

The Gram-positive teichoic acids

In the absence of an outer membrane, Gram-positive bacteria have developed alternative protective surface structures, such as capsular polysaccharides, S-layer proteins or mycolic acids in Actinobacteria. In addition to these variable features a common feature of all Gram-positive cell envelopes is the presence of teichoic acids (TAs). TAs were discovered in the 1950s by James Baddiley and coworkers. Two types

of teichoic acids can be distinguished: the wall teichoic acids (WTAs), covalently linked to the peptidoglycan layer through its N-acetylmuramic acid, and the lipoteichoic acids (LTAs) that are anchored to the cytoplasmic membrane by a glycolipid (Baddiley, 1972) (figure 1.2). These glycopolymers differ according to their net charge, to the size, redox-state and configuration of the oses involved in the linker and repeating units, as well as to the decoration of the repeating units (for a review on the structural diversity of TAs see (Weidenmaier and Peschel, 2008)). TAs often form a polyanionic network that provides the envelope with key physicochemical surface properties with respect to elasticity, porosity, strength and electrostatic steering which may largely impact biofilm formation for example. They also play a crucial role in bacterial physiology, in controlling susceptibility and/or resistance to antimicrobial peptides, and in maintaining cation homeostasis (Neuhaus and Baddiley, 2003). Located at the cell surface, TAs are involved in host cell attachment and environment colonization, and are also thought to activate pathways of the immune system in the case of bacterial pathogens (Weidenmaier and Peschel, 2008).

Some additional cell surface structures

To control motility, sporulation, local chemotaxis and host infections, bacteria have assembled large membrane embedded complexes. Swarming, twitching, gliding of bacteria on surfaces are finely tuned by different molecular apparatus organized in rotating flagella or retracting pili. For example, a rotor-mounted protein assembly, termed the "switch complex", formed from multiple copies of the 3 proteins FliG, FliM, and FliN, controls the direction of the flagellar rotation in E. coli. A combination of electron microscopy and high-resolution crystal structures have allowed the team of D. Blair to propose a mechanical description of the switch through protein reorientation (Paul *et al.*, 2011). Due to the relatively permeability of the outer membrane in Gram-negative bacteria, these organisms developed dedicated secretion systems to communicate (bioflim formation, slime layers), to inject toxins or effector proteins in host cells, or to defend against eukaryotic or bacterial predators. To date, seven secretion systems have been identified in bacteria. Among those the type III secretion system has been quite extensively characterized and the regulation of the interaction between over 20 cytosolic, periplasmic, and membrane-embedded proteins has been partially deciphered (Izoré *et al.*, 2011). In certain bacteria the cell surface is decorated with either a well-organized polysaccharide layer (the capsule) or a protein layer (the S-layer). Both structures are considered to play a role in virulence or in defense mechanisms.

The peptidoglycan

Peptidoglycan plays a central role in the maintenance of bacterial shape, prevention of cellular stress due to differences in osmotic pressure, and formation of daughter cells during the cell division process. The peptidoglycan is a complex heteropolymer of sugar and peptide chains, resulting from the polymerization and the incorporation to the preexisting peptidoglycan of a disaccharide-pentapeptide precursor (see Chapter 2 for more details). Its chemical composition is very similar in both Gram-positive and Gram-negative bacteria, although it differs by the thickness. As the absence of peptidoglycan strongly weakens the cellular integrity of bacteria, leading in most of the cases to cell lysis or death, its biosynthesis is the favorite target of some antibiotic classes.

1.2 Description of the main antibiotic families and their targets

The term "antibiotic" designates chemical compounds that, at low concentration, can prevent the growth of microorganisms (bacteriostatic antibiotic) or kill them (bactericidal antibiotic). Antibiotics specifically target and block one of the essential steps of the bacterial development: cell wall biosynthesis, protein or DNA synthesis. The first antibacterial agents used for human care in 1937 were the sulfonamides, also referred to as sulfa drugs. The first antibiotic, known as penicillin G, was in fact discovered in 1928 by Sir Alexander Fleming, but was only introduced for therapeutic use in 1941 during World War II. The use of antibiotics in medicine has been one of the major therapeutic progress of XX^{th} century, as it extended the life expectancy by more than 10 years in developed countries. However, the generalized and abusive use of antibiotics introduced a selective pressure, leading to the development of resistant bacterial populations. In this chapter, the main antibiotic families and their potential targets will be introduced.

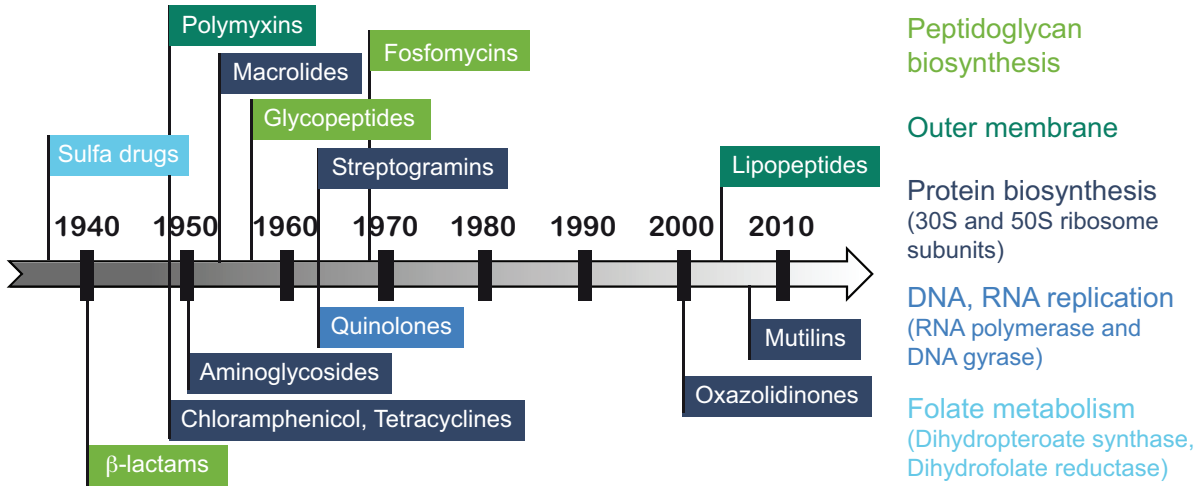


Figure 1.3: Antibiotics discovery along the years. Discovery success was approximately constant from 1936 to 1969. Then, a gap of 31 years was observed before the discovery of oxazolidinones. The colors correspond to the five categories of antibiotics targets. Data are taken from (Walsh, 2003) and (Norrby *et al.*, 2005).

Antibiotics can be divided into two main classes: the broad-extended-spectrum antibiotics which act on both Gram-positive and Gram-negative bacteria, and the narrow-spectrum antibiotics which are effective against either Gram-positive or Gram-negative bacteria. Furthermore, more than eighty years after the discovery of penicillin, a lot of progress has been made and antibiotic discovery efforts can be divided into two main streams: first, the discovery of the antibacterial activity of natural products, and a second breath which came from synthetic chemistry (Walsh, 2003). A large number of antibiotics are natural products, directly produced by microorganisms (bacteria and mushrooms). The discovery of this natural antibacterial activity permitted the development of a large number of broadly used antibiotics, such as penicillins, cephalosporins, and vancomycin. From 1940 to 1960, this strategy was predominant (figure 1.3). Advantages of natural antibacterials reside in the fact that they generally have complex architectural scaffolds and dense functional groups that are useful for the specific interaction with targets in pathogenic bacteria. However, today the number of natural antibiotics used in human therapy is de-

creasing, because of the lack of availability, side effects, and resistance phenomena. A large number of compounds used in medicine thus comes from synthetic chemistry, where antibacterial agents can either be derived from natural antibiotics, or can come from structures that are not found in nature. To date, two groups of synthetic antibacterials have been developed: the sulfa drugs and the quinolones. However, despite the development of synthetic products, no novel antibiotic class has been discovered between 1969 and 2000, creating a 31-years-gap (figure 1.3) during which bacteria had time to develop resistance mechanisms.

Each class of antibiotics acts on a specific bacterial molecular component or enzymatic reaction. The main antibiotics targets are the peptidoglycan, the outer membrane, the protein synthesis (via the ribosome), the DNA/RNA replication, and the synthesis of folic acid (figure 1.4).

Target 1: Peptidoglycan biosynthesis

β-lactam antibiotics target the Penicillin-Binding Proteins (PBPs), enzymes involved in the transpeptidation, an essential step of the peptidoglycan biosynthesis. β-lactams are naturally produced by some bacterial strains, but today most of them are semi- or totally-synthetic molecules. At first, β-lactam antibiotics were selectively active against Gram-positive bacteria, but with the recent development in β-lactams synthesis, broad-spectrum antibiotics in this family have been developed, ampicillin being one important example. The β-lactam family is to date the most broadly used antibiotic family as it represents more than 65 % of the world antibiotics market (Poole, 2004). More details on β-lactams will be found in the section 2.3.3.1 devoted to this class of molecules.

Fosfomycin inhibits bacterial cell wall biogenesis by inactivating the MurA enzyme, involved in peptidoglycan biosynthesis. This enzyme catalyzes the transfer of the *enol*pyruvate moiety of Phospho*enol*-pyruvate (PEP) to the 3'-hydroxyl of Uridine DiPhosphate (UDP)-*N*-acetyl-glucosamine (GlcNAc) (figure 2.5 page 44). As fosfomycin is a PEP analog, it inhibits MurA by creating a covalent bond with its active site cysteine residue (Dai *et al.*, 2002). Fosfomycin has a broad antibacterial activity against both Gram-positive and Gram-negative pathogens, with useful activity against *Enterococcus faecalis* (*E. faecalis*), *E. coli*, and various Gram-negative bacteria such as the *Citrobacter* and *Proteus* genera.

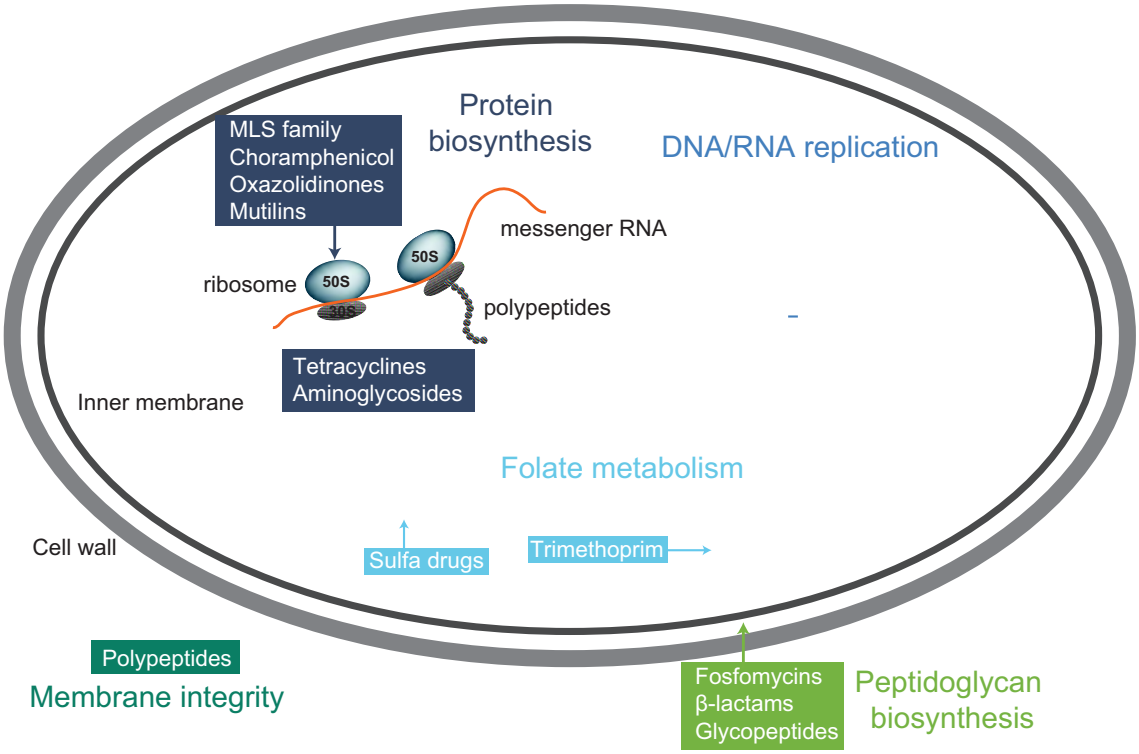
Finally, **glycopeptides** belong to the third class of antibiotics that targets the peptidoglycan. Their action mechanism involves the inhibition of extracytoplasmic steps in the peptidoglycan biosynthesis. Specifically, glycopeptide antibiotics interact non-covalently with the D-Ala-D-Ala terminus of the uncross-linked peptidoglycan pentapeptide (Pootoolal *et al.*, 2002). Glycopeptides are narrow-spectrum antibiotics, but they are crucial for the treatment of serious infections caused by Gram-positive bacteria, especially *Staphylococci, Streptococci, Enterococci*, and *Clostridia*. The two most broadly used glycopeptides are vancomycin and teicoplanin.

Target 2: The plasma and outer membrane

Antibiotics targeting the membrane are polypeptidic antibiotics.

First, the **lipopeptides** are very effective in fighting multidrug-resistant Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* (*S. aureus*) (MRSA) and vancomycin-resistant enterococci. A novel lipopeptide antibiotic is the daptomycin, whose mechanism involves insertion of the lipophilic antibiotic tail into the bacterial plasma membrane, causing rapid membrane depolarization and a potassium ion efflux. This loss of membrane potential leads to inhibition of DNA, RNA and protein synthesis, resulting in bacterial cell death (Steenbergen *et al.*, 2005).

The **polymyxins**, produced by the Gram-positive *Bacillus polymyxa*, are narrow-spectrum antibiotics active on Gram-negative bacilli. Polymyxins are surface-acting ampipathic agents. Each polymyxin molecule has a cationic polypeptide ring with a lipophilic fatty acid side-chain. The polypeptide ring binds with the anionic phosphate moieties of the bacterial outer membrane, displacing the cations which are needed for membrane integrity. This results in increased permeability of the cell membrane causing leakage of cellular contents and leading to cell death. The disruption of membrane integrity also increases the susceptibility of the organism to hydrophilic antibiotics such as rifampicin, carbapenems, glycopeptides and tetracyclines, thus opening the way for both Gram-negative and Gram-positive antimicrobial cooperative combination therapy (Balaji *et al.*, 2011).



with protein translation. The binding site of tetracyclines appears to be in the region of the transfer RiboNucleic Acids (tRNA) acceptor site, generally known as A-site. This binding sterically blocks aminoacyl-tRNA binding and results in protein synthesis inhibition (Thaker *et al.*, 2010), leading to the arrest of cell growth.

Chloramphenicol, also known as chlornitromycin, is very similar to tetracyclines. It is also a bacteriostatic antimicrobial, and is effective against a wide variety of Gram-positive and Gram-negative bacteria, including most anaerobic organisms. Chloramphenicol inhibits the peptidyl transferase activity of the bacterial ribosome, thus preventing the protein chain elongation.

Aminoglycosides are potent bactericidal antibiotics targeting the bacterial ribosome, where they bind to the A-site and disrupt protein synthesis. They are particularly active against aerobic, Gram-negative bacteria and act synergistically against certain Gram-positive organisms. Bacterial resistance to aminoglycosides continues to increase and is widely recognized as a serious health threat (Jana and Deb, 2006). Streptomycin was the first of this class to be discovered in 1943, and was the first antibiotic remedy for tuberculosis.

Another class of antibiotics targeting the ribosome is the class of **macrolides**. The broad spectrum activity of macrolides against the Gram-positive and Gram-negative respiratory pathogens has led to their widespread acceptance for the treatment of community-acquired respiratory pathogens, including pneumonia, pharyngitis, and tonsillitis (Pechère, 2001). Macrolides inhibit bacterial growth by binding to the ribosome and blocking protein synthesis. They bind in the nascent peptide exit tunnel and disturb the passage of the newly synthesized polypeptides through the tunnel, thereby interrupting translation elongation (Kannan and Mankin, 2011).

Lincosamides prevent bacteria replication by interfering with the synthesis of proteins in sensitive microorganisms. The antibacterial activity of lincosamides relies on the inhibition of the 23S portion of the 50S subunit of bacterial ribosomes and cause premature dissociation of the peptidyl-transfer-RiboNucleic Acid (tRNA) from the ribosome. Clindamycin is the main lincosamide antibiotic that is used in clinical practice (Spízek *et al.*, 2004).

Streptogramins are a group of natural cyclic peptides produced by several Streptomyces species. Both group A and group B streptogramins bind to bacterial ribosomes and inhibit the translation of messenger RNA (mRNA) at the elongation step (Khosla *et al.*, 1999).

Macrolide, lincosamide, and streptogramin form the so-called “MLS” family. Even if they are chemically distinct, they have a similar mode of action (Leclercq and Courvalin, 1991; Tenson *et al.*, 2003).

Oxazolidinones are targeting a step involving the binding of N-formyl-methionyl-tRNA to the ribosome, and act thus as protein synthesis inhibitors. One example of recent clinically used oxazolidinone is the linezolid, which was approved in 2000 and became the reference of this antibiotic class (Shaw and Barbachyn, 2011). Some of the most important oxazolidinones are the last generation of antibiotics used against gram-positive pathogens, including methicillin-resistant *S. aureus*.

Finally, **mutilins** (or pleuromutilins) are antibacterial drugs that inhibit protein synthesis in bacteria by binding to the peptidyl transferase component of the 50S subunit of ribosomes (Long *et al.*, 2006). Mutilin derivatives, which were successfully developed in veterinary medicines (valnemulin and tiamulin for example), have regained interest as promising antibiotics with potential for human use. In 2007, retapamulin was one of the rare new class of antibacterial approved in the last decades for use in human skin infections.

Target 4: DNA and RNA replication

The **quinolones** and **fluoroquinolones** target two essential bacterial enzymes: the DNA gyrase (topoisomerase II) and the DNA topoisomerase IV. These two enzymes act by passing one region of duplex DNA through another. During that process, the quinolones form complexes with enzyme and DNA (Drlica *et al.*, 2009). The complexes block movement of replication forks and transcription complexes, thereby inhibiting bacterial growth. Quinolones display broad spectrum activity, in particular with *pseudomonas* and *staphylococcus*. Largely used quinolones are nalidixic acid, ciprofloxacin and many others.

Rifamycins belong to a group of antibiotics that are either naturally produced by the bacterium *Amycolatopsis mediterranei* or synthetically created. The rifamycins display a broad spectrum of antibiotic activity against Gram-positive and, to a lesser extent, Gram-negative bacteria. Rifamycins are particularly effective against mycobacteria, the group member rifampicin has thus become one of the main agents used in the treatment of tuberculosis, leprosy, and associated mycobacterial infections. The mechanism of action of rifamycins is via inhibition of bacterial DNA-dependent RNA synthesis. This is driven by the high affinity of rifamycins to the DNA-dependent RNA polymerase of prokaryotes (Floss and Yu, 2005).

Target 5: Folic acid synthesis

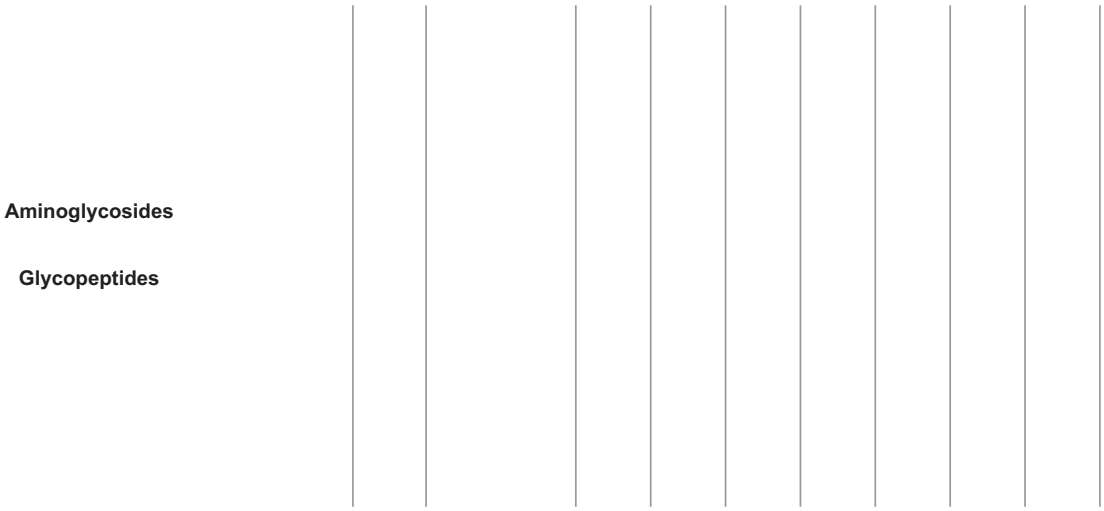
Sulfonamides are broad-spectrum, bacteriostatic antibiotics. Also called **sulfa drugs**, they were the first synthetic antibiotics. This family was discovered in 1936 by Domagk in a sulfur-containing dye that was able to cure mice infected by *streptococcus*. Thus, Prontosil, the first sulfa drug antibiotic, started to be used in treatment of human streptococcal infections in 1936 (Bentley, 2009). Then, the sulfanilamide, another derivative from dye, was shown to be an effective antibacterial agent. As it was easy to modify by organic chemistry, a lot of sulfa drugs have been synthesized and more than 5,000 new antibiotics of this class were generated by 1945 (Domagk, 1957). As an example, sulfapyridine, available in 1938, was extensively used in the treatment of pneumonia. The mechanism of action of sulfonamides is a competitive inhibition of the enzyme DiHydroPteroate Synthetase (DHPS), an enzyme involved in the folate biosynthetic pathway. However, because of associated toxicity and high rates of resistance, their use is now very limited.

1.3 Antibiotic resistance phenomenon

Despite the large number of antibiotics available and their various targets within bacteria, the increasing phenomenon of antibiotic resistance is worrying as it develops very fast, faster than our ability to design new inhibitors (Livermore, 2012). This section will deal with the emergence of resistance and the part of human role in this; then, the 4 main mechanisms of resistance that were developed by bacteria these last decades will be detailed.

1.3.1 Emergence of resistance

Since the beginning of antibiotic discovery and use, bacteria developed different mechanisms to resist against antibacterial agents. The accelerated emergence of bacteria that are resistant to multiple antibiotic types now appears as a very serious threat about the success in the treatment of infectious diseases in the future. It is also striking to note how fast bacteria develop resistance, as shown in figure 1.5. In general, resistance for most classes of antibiotics is typically reported within four years following their FDA approval. In some cases, resistance was already present in the environment before the agent was approved for clinical use (penicillin, linezolid). This rapid development of resistance has been largely stimulated by the systematic and sometimes unjustified use of bactericide molecules by man. Concurrently, the extensive use of antibiotics in animal breeding took part in the propagation of resistant strains. Some bacterial species, naturally resistant, have been slowly favored because of intensive antibiotics use on both human and animals. The selective pressure on all the microorganisms has increased, leading to some multi-resistant strains in hospital of bacteria resistant to all of the currently used and active antibiotics (De Lencastre and Tomasz, 2002). As an example, the extensive use of fluoroquinolones and other wide-spectrum antibiotics such as cephalosporins, has led to the rapid diffusion of *Clostridium difficile* Associated Diarrhea (CDAD), particularly in elderly and immunocompromised patients.



The expansion of resistance and multi-resistance is combined with the lack of new antibiotic classes developed by pharmaceutical companies. Only four new classes of antibiotics have reached the market in the last ten years: linezolid (oxazolidinone class), daptomycin (cyclic lipopeptide), retapamulin (mutilin class) and fidaxomicin (macrolactone), which were approved by the FDA in 2000, 2003, 2007 and 2011, respectively (Jabes, 2011). All the new drugs are targeting Multi-Drug Resistant (MDR) Gram-positive pathogens and CDAD. Despite the real need for new antibacterial agents to stop the increasing antibiotic resistance, pharmaceutical industries progressively decreased their investment in antibiotic research to devote themselves to more profitable domains (Livermore, 2012). Indeed, more than 90% of infectious can be cured by available molecules, and the antibiotic treatments are very short compared to treatments of other diseases, such as chronic diseases.

Research on new antibiotics remains nevertheless the only remaining option to fight against bacterial resistant strains, in combination with a public health strategy taking care of the non-abusive use of existing and future antibacterial agents. In order to broaden the possibilities of new molecules creation, the understanding of the original resistance mechanism is required, in addition to all potential targets and their location.

1.3.2 The 4 main mechanisms of resistance

Bacteria developed different mechanisms to efficiently fight antibiotics, β-lactams in particular. So far, four resistance mechanisms have been identified (see figure 1.6):

- The production of enzymes that degrade or modify the antibiotic before it can reach its target. β-lactamases are the main example of this class of enzymes (Wilke *et al.*, 2005)
- The alteration of the antibiotic target site, for example through the development of low affinity PBPs
- The prevention of antibiotic access to the target, for example through alteration of the efflux pump function
- The activation of an alternative cross-link pathway, as for example by enhancing a L,D-transpeptidase activity

β-lactamases

In Gram-negative pathogens, β-lactamase production is the most common mechanism involved in the β-lactam resistance phenomenon. These enzymes disrupt the amide bond of the β-lactam ring, rendering the antibiotic inactive (figure 1.6) (Majiduddin *et al.*, 2002; Helfand and Bonomo, 2003). β-lactamases were first identified in *S. aureus* in the 1940s, prior to the clinical use of penicillin (Abraham and Chain, 1940). Four classes of β-lactamases have been described, based on the primary sequence and the catalytic mechanism. Classes A, C, and D are enzymes with a serine in the active site that catalyzes hydrolysis of the β-lactam ring through an acyl-intermediate, whereas the class B enzymes, also referred to as Metallo-β-Lactamases (MBLs), require either a single or two Zn^{2+} ions (Ambler, 1980) coordinated to His/Cys/Asp residues in their active site. Currently, more than 850 β-lactamases are identified (Drawz and Bonomo, 2010), and the rapid replication and recombination rate, as well as the high mutation frequency, will probably permit bacteria to resist against novel β-lactams thanks to the evolution of β-lactamases.

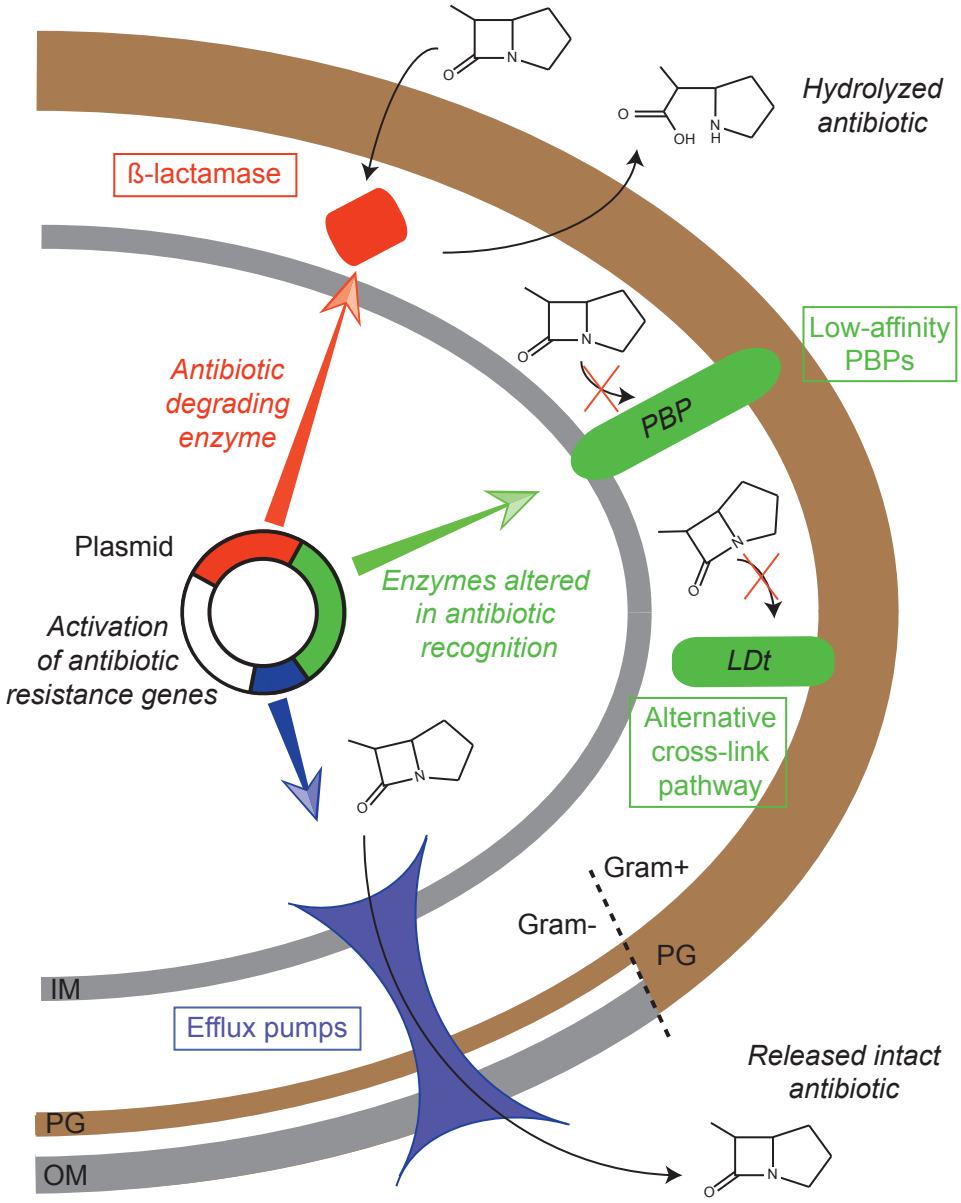


Figure 1.6: Mechanisms of antibiotic resistance in the bacterial cell wall. Efflux pumps are specific to Gram-negative bacteria, however the other resistance mechanisms can apply to both Gram-positive and Gram-negative bacteria. Abbreviations: Gram+/-: Gram positive and negative; IM: inner membrane; OM: outer membrane; PG: peptidoglycan.

The class A, C and D of β-lactamase enzymes share structural similarities with the D,D-peptidases. The first structure was solved in 1987 for the class A β-lactamase from *S. aureus* (Herzberg and Moult, 1987), and since then a large number of β-lactamase structures have been determined, the class A being the most studied group (Majiduddin *et al.*, 2002). These structures revealed four major catalytic elements in the three classes: the S-X-X-K, S-X-V and K-T/S-G motifs, as well as a loop referred to as Ω-loop (Leszczynski and Rose, 1986). The first motif includes the active serine, and, separated by two residues, a lysine that is essential for the acylation with β-lactams. More precisely, this lysine, Lys^{73} in *S. aureus*, participates in the proton transfer of the catalytic serine to the amide nitrogen of the β-lactam ring along the acylation reaction. The glutamate Glu^{166}, also in the active site, catalyzes the deacylation of the complex (Herzberg and Moult, 1987). In class B enzymes, it is the Tyr^{150} that is suggested to act as a proton acceptor for the OH group of the active serine, which would then be activated for the nucleophilic attack of the carbonyl of the β-lactam ring. In class D enzymes, the serine would be activated thanks to Lys^{70}, the structural equivalent of Lys^{73} in class A enzymes, via the initial activation of a water molecule.
β-lactamases of class A are particularly implicated in the problem of resistance to β-lactam antibiotics, as the β-lactamase genes can be transfered within and between species. This dissemination of the genes occurs via plasmids and transposons. As an example, the TEM-1 β-lactamase, a major source of resistance in Gram-negative bacteria, which was first identified in *E. coli*, has then been acquired by other pathogens (Frère, 1995). Furthermore, the extensive clinical use of extended-spectrum antimicrobial agents has resulted in point mutations in existing β-lactamases, providing them with a large resistance against extended-spectrum β-lactams. The mutations occur near the active site of the enzymes and increase the catalytic efficiency for hydrolysis of antibiotics (Huang *et al.*, 1996).

The Metallo-β-Lactamases (MBLs) are further divided into three functional groups based on metal requirements: B1, B2 and B3 (Majiduddin *et al.*, 2002). Despite a low sequence homology, the overall fold of these enzymes is very similar, as they all share an αβ/βα-sandwich structure comprising two β-sheets at the core and five α-helices on the external faces (Wang *et al.*, 1999). MBLs have been identified in both Gram-positive and Gram-negative bacteria, and their coding genes are often found on integron elements, which renders the spread easier (Lauretti *et al.*, 1999; Laraki *et al.*, 1999). These metalloenzymes represent a particular clinical threat because, unlike the serine β-lactamases (class A, C and D), they are not inactivated by most of the β-lactam antibiotics. Furthermore, they are often expressed in combination with other β-lactamases and have a broad substrate profile, enabling them to inactivate a large panel of antibiotics.

Development of low affinity PBPs

PBPs represent the lethal targets of penicillins and more generally of β-lactam antibiotics. The β-lactam ring exhibits structural analogy with the D-alanyl-D-alanine of peptidoglycan pentapeptides, the natural substrate of PBPs (see chapter 2). β-lactam antibiotics acylate the penicillin-binding domain of PBPs, forming a stable acylenzyme, then destitute of its transpeptidase activity. However, some bacterial strains are naturally resistant to this class of antibiotics. This was correlated to the presence of point mutations in some resistant strains, as in PBP2 from *Streptococcus pneumoniae* (*S. pneumoniae*) where a succession of seven mutations has been directly linked to penicillin resistance (Hakenbeck and Coyette, 1998). These mutations were located between the S-X-N catalytic triad motif and the active serine of the

enzyme.

Enterococci species are particularly resistant to β-lactams (10 to 1000 times more than streptococci). This natural resistance is linked to the production of a PBP with low affinity for this class of antibiotics (Macheboeuf *et al.*, 2006). In *Enterococcus faecium* (*E. faecium*), two mechanisms are believed to be involved in the high-level resistance to benzylpenicillin: either these strains overproduce a large quantity of the PBP5$_{fm}$ with the same affinity for the antibiotic (Fontana *et al.*, 1983, 1994), or they produce a PBP5$_{fm}$ with an apparent decreased affinity for benzylpenicillin (Klare *et al.*, 1992). The structure of PBP5$_{fm}$ has been modelized (Zorzi *et al.*, 1996) and then solved (Sauvage *et al.*, 2002) in order to understand where the natural resistance could originate from. The X-ray structure consists of a classical transpeptidase domain preceded by an elongated N-terminal region. A residue has been identified as responsible for a 16-fold increase in the Minimum Inhibitory Concentration (MIC) values upon mutation: the Met^{485}, located three residues after the S^{480}-X-N^{482} catalytic triad motif. It was suggested that the absence of the long methionine side-chain could allow the Lys^{425} to have more space and to thus be less active in the acylation mechanism by benzylpenicillin (Sauvage *et al.*, 2002).

Efflux pumps

Gram-negative bacteria possess a supplementary system to resist against antibiotic attacks: the efflux pumps. These transport proteins prevent the entrance or favor the extrusion of antibiotics from within the cells to the external environment (figure 1.6). Nowadays, efflux pumps contribute significantly to bacterial resistance as they recognize a broad variety of substrates. They are expressed in important pathogens and cooperate with other resistance mechanisms (van Bambeke *et al.*, 2000). Efflux pumps are active transporters, as they require an external source of chemical energy to perform their function. Some are primary active transporters utilizing Adenosine TriPhosphate (ATP) hydrolysis as a source of energy (example: The ATP-binding cassette family (ABC)), whereas others are secondary active transporters (uniports, symports, or antiports) in which transport utilizes the transmembrane proton gradient as the driving force for multidrug efflux (Paulsen *et al.*, 1996). Within this class of secondary active transporters, five superfamilies have been described, based on the amino acid sequence:

- The Small Multidrug Resistance (SMR)
- The Major Facilitator Superfamily (MFS)
- The Multi Antimicrobial Resistance (MAR)
- The Resistance-Nodulation Division (RND)
- The Multidrug Endosomal Transporter (MET)

Genes encoding efflux pumps can be found on plasmids and are carried on the chromosome, which provides bacteria an intrinsic resistance to antibiotics (Webber and Piddock, 2003), especially in mutant bacteria which over-express these genes. Another characteristic of efflux pumps is their skill to transport a large variety of molecules, as they exhibit a poor substrate specificity. Most transporters can recognize molecules with a polar and slightly charged domain, combined with a hydrophobic tail like fluoroquinolones, β-lactams, chloramphenicol and trimethoprim (van Bambeke *et al.*, 2000). Only antibiotics with hydrophilic properties such as glycopeptides, that are not able to cross the bacterial inner membrane and that show a surface activity, are not susceptible to efflux-pump type of resistance mechanism.

An alternative peptidoglycan cross-link pathway

Peptidoglycan biosynthesis is one of the main target of antibiotics. The transpeptidation mechanism, that cross-links the glycan strands and peptide stems together to form the peptidoglycan, is thus especially targeted. Transpeptidation is usually performed through a D,D-transpeptidation, however, under antibiotic stress, bacteria can activate an alternative pathway involving L,D-transpeptidation. This novel pathway enables bacteria to neglect the presence of antibiotics. This resistance phenomenon is the one we will focus on in this thesis, and will thus be detailed in the section devoted to L,D-transpeptidases.

Despite the large variety of antibiotics and the variability of the bacterial targets investigated, either cytoplasmic or on the cell surface, bacteria have found a way to deviate the antibiotic treatments. Among the targets, peptidoglycan remains one of the principal ones due to its relative accessibility both in Gram-positive and Gram-negative bacteria. With the emergence of alternative pathways to polymerize the elementary units and the potency to develop new approaches to inhibit these steps, a renewed interest towards this bacterial component is emerging. It is thus the focus of the next chapter.

Chapter 2

The peptidoglycan

2.1 Peptidoglycan physicochemical characteristics

2.1.1 Peptidoglycan chemical composition

The disaccharide pentapeptide monomer

The primary structure of peptidoglycan consists of linear glycan strands which are cross-linked together by short peptides (Rogers *et al.*, 1980). The glycan strands are composed of alternating *N*-acetylglucosamine (GlcNAc) and *N*-Acetyl-Muramic Acid (MurNAc) residues, linked together by β-1,4 glycosidic linkages, as shown in figure 2.1 for the peptidoglycan of *E. coli*.

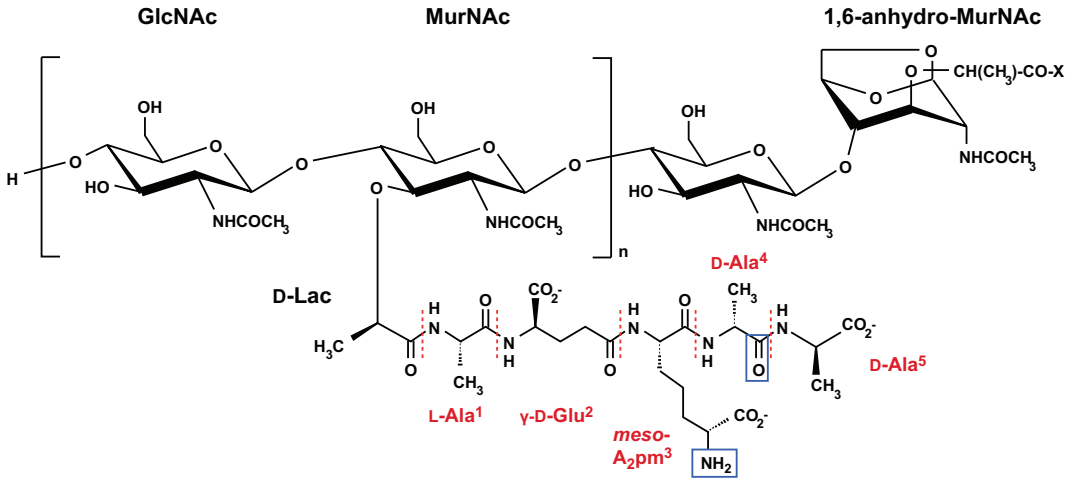


Figure 2.1: Repeating unit of *E. coli* peptidoglycan. Glycan strands consist in alternating β-1,4-linked GlcNAc and MurNAc residues, and are terminated by a 1,6-anhydroMurNAc as in all Gram-negative bacteria. D-Lac represents the lactyl group of MurNAc and is the anchor of the peptide stems to the glycan strands. The five residues of the dissacharide pentapeptide are separated by dashed red lines. X indicates a peptide stem bound to the D-Lac of the ending MurNAc. The blue frames around the NH_2 of *meso*-A_2pm and the carbonyl of D-Ala^4 indicate the anchor for the cross-link between the acceptor stem and the donor of a second peptide stem (devoid of D-Ala^5), respectively.

Pentapeptide stems are attached by an amide linkage to the lactyl group (D-Lac) of each MurNAc residue. In most Gram-negative bacteria, the composition of the pentapeptide attached to the dissacharide is L-Ala^1-γ-D-Glu^2-*meso*-A_2pm^3-D-Ala^4-D-Ala^5. In Gram-positive bacteria the *meso*-A_2pm is generally replaced by L-Lys (see Table 2.2 on page 47). This ensemble forms the basic peptidoglycan subunit, commonly called the disaccharide pentapeptide motif. Glycan strands are then cross-linked through the stem peptides to form a net-like structure (Glauner *et al.*, 1988) and are terminated with a 1,6-anhydro-MurNAc instead of a simple MurNAc in some species. Cross-linking generally occurs between the carbonyl group of the D-Ala^4 of the first peptide stem (donor) and the amino group of the amino acid at

position 3 of the second peptide stem (acceptor). This cross-link is accordingly named the 4→3 cross-link. In the meantime, the D-Ala^5 is released along the cross-linking reaction.

Characteristic features of the polymer

The first questions raised after the identification of the repetitive motif in a polymer concern the degree of polymerization (to estimate the average number of units present in a typical chain) and the polydispersity (distribution of this number). In the case of the peptidoglycan heteropolymer, these questions can address either the length of the glycan strands or the degree of cross-linking of the peptide stems.

The mean length of the glycan strands in peptidoglycan can be determined by two different approaches (Vollmer *et al.*, 2008):

- the quantification of the ending sugars (by chemical reduction, by addition of radiolabeled galactosamine to the GlcNAc termini or by identification of the 1,6-anhydroMurNAc termini) with respect to the total sugar content of the polymeric material;
- the purification, separation by High-Performance Liquid Chromatography (HPLC) and identification by mass-spectrometry of glycan strands released after amidase treatment.

The average glycan chain length GCL has been estimated to fall between 20 and 40 disaccharide units in most Gram-negative bacteria, the exact values depending on growth conditions and stages (Vollmer and Seligman, 2010). In particular, studies in *E. coli* revealed that GCL was varying depending on the stage in cell growth and suggested a large and disperse GCL distribution. In exponential growth, *E. coli* strains have average GCL comprised between 25 and 35 disaccharide units, whereas this number drops to around 15 disaccharide units in resting cells and increases to 50-60 in newly synthesized material (Table 2.1, (Pisabarro *et al.*, 1985; Glauner, 1988; Glauner and Höltje, 1990)). Average value and distribution of GCLs in Gram-positive organisms appears to be even more disperse and species dependent. As an example, glycan strands of up to 5,000 disaccharide units were observed in *B. subtilis* (with an average size of more than 500 units) (Hayhurst *et al.*, 2008), whereas 85-90 % of the strands in *S. aureus* consist of glycan strands with a mean length of 6 disaccharide units only (Boneca *et al.*, 2000).

The degree of peptide cross-linking is determined after separation by HPLC of the fragments obtained after peptidoglycan digestion by mutanolysine or lysozyme (these enzymes cleave the β-1,4-glycosidic linkage between MurNAc and GlcNAc in this order). Each fragment is then identified by mass-spectrometry and the percentage of cross-linking is established after quantification of each fragment (Glauner, 1988; Magnet *et al.*, 2007a). Alternative solid-state methods are emerging to quantify the degree of cross-linking. Their advantage resides in the analysis of intact peptidoglycan instead of digested material. Their limitation stands in the requirement for ^{13}C and or ^{15}N isotope labeling. Specific labeling is required in 1D strategies promoted by the group of J. Schaefer (Patti *et al.*, 2008) and uniform labeling is required in the strategies developed in the group of Jean-Pierre Simorre, at the IBS (Gansmüller et al., manuscript in preparation). There is a considerable variation in the degree of cross-linking, which varies from ca. 20 % in *E. coli* to over 93 % in *S. aureus* (Table 2.1). Taking into account these numbers, Gram-negative bacteria, with the thinner peptidoglycan layer, would on average have a cross-link every second to third disaccharide units, while the highly cross-linked *S. aureus* would contain 10-100 nm long peptide stems linked to multiple short-length glycan strands. Characteristic patterns

described here concern mainly average values at a close-to-atomic resolution. They may not reflect the polydispersity observed at a larger scale for biophysical properties of the material.

Organism (growth conditions), remarks	Mean length of glycan strands		Cross-linkage (% [c])
	DS units [a]	Method	
Gram-negatives			
Escherichia coli KN126 (LB medium, exponential growth)	25.8	Anh [b]	48.5
E. coli KN126 (LB medium, stationary)	17.8	Anh	60.6
E. coli KN126 (PB medium, 30 °C)	33.3	Anh	44.7
E. coli KN126 (PB medium, 42 °C)	37.9	Anh	44.6
E. coli W7, different growth rates	23–30	Anh	45 – 49
E. coli W7, newly made peptidoglycan	50–60	Anh	∼40
E. coli EH3247	8.9 (70 –75%)	GL [d]	
	45.1 (25 –30%)	Anh	
	21 (total)	GL/Anh	
E. coli P678-54, isolated minicells (cell poles)	23.5	Anh	54.8
E. coli P678-54, whole cells	27.8	Anh	52.4
E. coli MC4100 lysA , normal rod-shaped	31	Anh	47
E. coli MC4100 lysA ftsA , filamentous cells	30	Anh	46
E. coli MC4100 lysA pbpB , spherical cells	32	Anh	47
E. coli , different strains and growth conditions	33–52	Gal [e]	
E. coli MC6RP1	22	Anh	30.8
Acinetobacter acetoaceticus	11	Anh	49.0
Aeromonas sp.	36	Anh	32.5
Agrobacterium tumefaciens	38	Anh	32.7
Enterococcus cloacae ATCC 13047	31	Anh	35.6
Helicobacter pylori , NCTC 11637, spiral cells (4 h of growth)	7.1	Anh	27
H. pylori NCTC 11637, round cells (stationary)	5.6	Anh	31
H. pylori 26695 (8 h of growth)	9.1	Anh	29.3
H. pylori 26695 (48 h of growth)	7.9	Anh	31.1
Myxococcus xanthus	9	Anh	66.3
Proteus morganii	53	Anh	36.1
Pseudomonas aeruginosa	16	Anh	34.6
Pseudomonas putida	29	Anh	25.2
Thermus thermophilus HB8	30	Gal	25
Vibrio parahaemolyticus	10	Anh	31.9
Yersinia enterocolitica	36	Anh	29.2
Gram-positives			
Bacillus subtilis	> 500	AFM [f]	
B. subtilis HR (growing)			56.4
B. subtilis HR (late stationary)			63.9
Deinococcus radiodurans	20	Anh	39.5
Staphylococcus aureus , different strains	∼6 (85 –90%)	GL	74 –92
	> 26 (10 –15%)		

[a] DS: disaccharide. One DS unit has a length of approximately 1.03 nm.
[b] Anh: From the analysis of muropeptide composition; calculated form the fraction of 1,6-anhydroMur N Ac-containing muropeptides (MurNAc ends).
[c] Percentage of peptides present in cross-links, from the analysis of muropeptide composition; calculated as 100% –% monomeric muropeptides.
[d] GL: HPLC separation and quantification of isolated glycan strands.
[e] Gal: quantification of Glc N Ac ends by enzymatic attachment of radioactive D-Gal.
[f] AFM: atomic force microscopy.

Table 2.1: Glycan chain length and cross linking of peptidoglycan from different species. Taken from (Vollmer and Seligman, 2010).

2.1.2 Biophysical properties of the peptidoglycan

The ultrastructure of the peptidoglycan polymer is characterized at the same time by an important rigidity and an impressive elasticity. As an example, the *E. coli* sacculus can sustain a 25 atm pressure. This rigidity is due to the glycan chains, as their rotation is limited by the β-1,4 inter-saccharide linkage. The elasticity is conferred by the peptides, which enables a 3-fold narrowing or expansion without rupture of the cell wall. The *E. coli* sacculus elasticity is greater along the long axis with an average elastic modulus of 2.5×10^7 N.m^{-2} than in the direction perpendicular to the long axis (4.5×10^7 N.m^{-2}) (Yao *et al.*, 1999). Solid-state Nuclear Magnetic Resonance (ssNMR) studies in the team of Jean-Pierre Simorre at the IBS also confirm the relative increase of the flexibility of the peptides with respect to that of the glycan chains (Kern *et al.*, 2008). Cryo-transmission electron microscopy (cryo-TEM) has recently offered key information on the thickness of the peptidoglycan layer in different Gram-positive and Gram-negative species that are in agreement with small angle diffusion light scattering (SANS) data and atomic force microscopy (AFM). These data suggest that the peptidoglycan layer might not be as homogeneous as expected. As a matter of fact, according to SANS, the thickness of the hydrated peptidoglycan *E. coli* is 2.5 nm for 75-80 % of the polymer and 7 nm on the rest, suggesting a monolayer of peptidoglycan on most of the surface and a triple-layer of polymer in some regions (Labischinski *et al.*, 1991). Furthermore in Gram-positive, there is no direct correlation between the glycan chain length or the peptide cross-linking and the peptidoglycan thickness. In addition to irregularities of the macroscopic features, the peptidoglycan present pores with a radius of 2.0 to 2.5 nm depending of the bacterial species that can accommodate proteins up to 50 kDa. Despite the amount of chemical and structural information collected on the peptidoglycan and the large number of techniques used for its investigation, the three dimensional architecture of this biopolymer still escapes experimental determination with an atomic resolution. While the previously cited techniques do not offer the required resolution, X-ray crystallography could not be applied due to the non-crystallinity of the material. Solid-state NMR yielded only sparse information, providing few internuclear distances (Sharif *et al.*, 2009; Bougault *et al.*, 2012) that were insufficient to build a complete model. Important questions remain opened concerning the degree of order of the biopolymer, the orientation of the glycan strands, the conformation of the disaccharide peptide motif, and the organization of the different peptidoglycan layers in Gram-positive bacteria. In the absence of answers, different models have been proposed along the years to satisfy the available data, and are exposed below.

2.1.3 Peptidoglycan architecture: the proposed structural models

Three principal models have been proposed in the litterature. In the “layered model” (figure 2.2a), the glycan chains and peptides are arranged parallel to the membrane, whereas in the “scaffold model” (figure 2.2b), the glycan chains extend perpendicular to the membrane. Finally, the more recent “coiled-coiled model” (figure 2.4) proposes a more elaborate peptidoglycan architecture in *B. subtilis* with bands made of coiled 50 nm-wide bundles of glycan strands.

Layered model

The layered model was first suggested (figure 2.2a), and was based on its similarity to chitin, which is the main component of the cell walls of fungi and of the exoskeletons of arthropods such as crustaceans

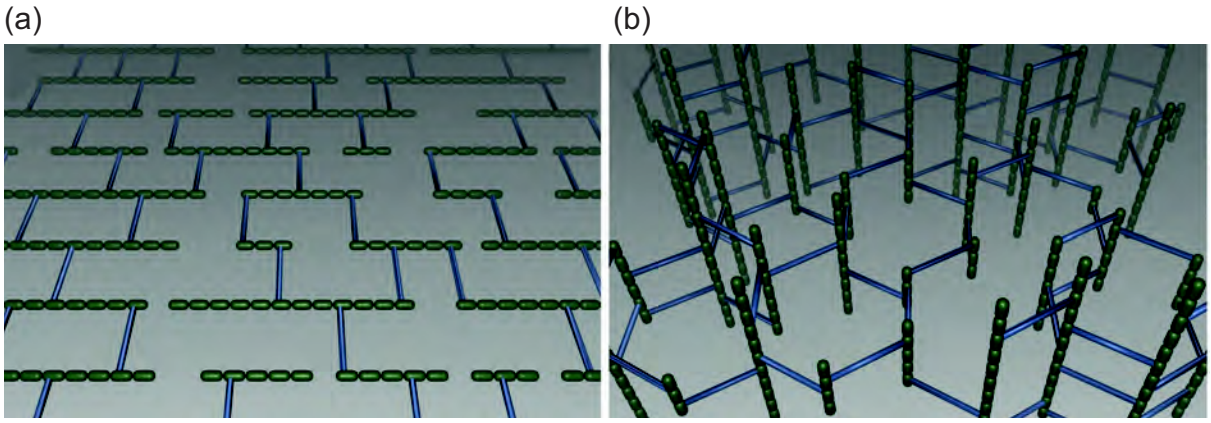


Figure 2.2: Representation of peptidoglycan a) layered and b) scaffold models, viewed obliquely to the sacculus surface. In the layered model, glycan strands are oriented parallel to the membrane. The cross-linked peptide stems are in the same plane as glycan strands, whereas the non-cross-linked peptides point out of the layer (non represented). In the scaffold arrangement, glycan strands are extended perpendicularly to the cytoplasmic membrane. Green pills represent the disaccharides (GlcNAc and MurNAc molecules), and blue lines represent the peptides. Figure taken from (Gan *et al.*, 2008).

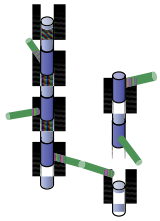
and insects. The glycan chains of peptidoglycan present similarity to chitin as the latter are made of GlcNAc oligomers bound through a β-1,4-linkage. First, all molecular models designed were based on the assumption that the arrangement of the glycan chains in peptidoglycan was the same as in chitin. Yet, in chitin, the successive GlcNAc residues are rotated by 180° relatively to each other (on a 2-fold axis, figure 2.3), and it was demonstrated using small and medium-angle X-ray scattering that the peptidoglycan can not adopt a chitin-like structure as the presence of the MurNAc residue does not allow such rotation (Labischinski *et al.*, 1979). Instead, four disaccharide units are needed to complete one turn (on a 4-fold axis, figure 2.3), and consequently the peptides protrude in different directions, arranged in a helical pattern. Furthermore, it has been shown in the same study that the glycan chains lie preferably in the plane of the cell wall. This means that every second peptide lies in the same plane and can then be cross-linked to neighboring glycan strands. Therefore, 50 % of the peptides would point up or down of the plane, so only the other half can take part in cross-links. With this model, a perfect monolayer of peptidoglycan would contain 50% of cross-linked peptides, and 50% of non-cross-linked peptides. This is in agreement with the experimentally cross-linking percentage observed in *E. coli* at about 40 - 50 % (Glauner *et al.*, 1988; Glauner and Höltje, 1990). Furthermore, recent images of frozen and hydrated sacculi isolated from Gram-negative *E. coli* and *Caulobacter crescentus* species using cryo-TEM showed a single layer of peptidoglycan in which the glycan strands lie parallel to the membrane (Gan *et al.*, 2008).

Scaffold model

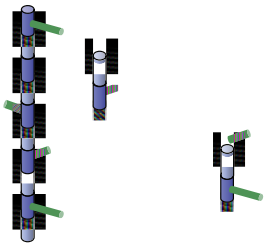
An alternative to a planar orientation of the glycan chain with respect to the membrane is the vertical orientation of glycan strands, forming a scaffold-like structure (figure 2.2b). This new model was introduced by Dmitriev *et al.* (Dmitriev *et al.*, 1999, 2000, 2003), as they did not manage to create a representation of the peptidoglycan architecture by computer simulation which was able to fit all of the known experimental data with the existing layered model. Thus, the authors decided to explore a completely different approach, using the method from Pink *et al.* (Pink *et al.*, 2000), with the possibility that

2-fold symmetry
(180°)

3-fold symmetry
(120°)



4-fold symmetry
(90°)



at least for *B. subtilis*. This model certainly needs to be supplemented by additional experiments, and molecular details should be determined in order to be validated and extended to other bacterial strains.

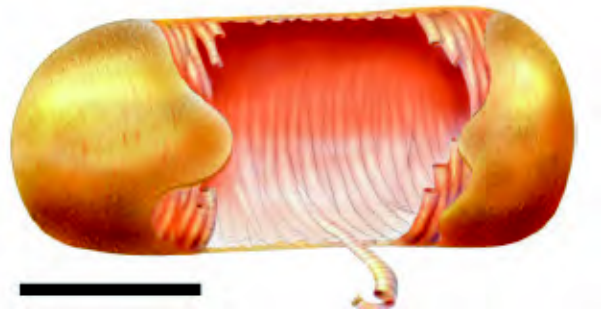


Figure 2.4: Model of cell wall architecture for *B. subtilis*, representing the coiled-coiled model of peptidoglycan arrangement. The cables of coiled substructures are running around the cylindrical cell envelope. The scale bar represents 1 µm. Figure taken from (Hayhurst *et al.*, 2008).

Validation and/or implementation of these models await additional experimental data. Due to the multiple layers of the peptidoglycan in Gram-positive bacteria, the situation may be even more complex than in Gram-negative organisms.

2.2 Peptidoglycan biosynthesis in the cytoplasm

Two cellular compartments are involved in peptidoglycan biosynthesis: the cytoplasm and the periplasm, through three main distinct stages (for a review, see (Lovering *et al.*, 2012)). The first stage involves reactions that occur within the cytoplasm, with the formation of UDP-GlcNAc, its transformation into UDP-MurNAc and the step-by-step assembly of amino acids onto the latter, to generate the UDP-MurNAc-pentapeptide. The second stage consists of the formation of Lipid I and Lipid II complexes from UDP-MurNAc-pentapeptide, and the transfer of Lipid II precursors from the cytoplasm to the periplasmic space. The third stage involves the polymerization of Lipid II in the periplasm to yield the peptidoglycan. The first two stages will be detailed in this section and are depicted in figure 2.5. The last stage of this biosynthesis will be detailed in the following section.

2.2.1 Synthesis of precursors

Biosynthesis of UDP-GlcNAc

The fructose-6-phosphate (Fructose-6-P) is the starting cytoplasmic material of this biosynthesis. It is a fructose sugar phosphorylated on carbon 6, whose β-D-form is very common in cells. The biosynthesis of uridine diphosphate-acetylglucosamine (UDP-GlcNAc) requires the modification of the Fructose-6-P by three successive enzymes: GlmS, GlmM, and the bifunctional enzyme GlmU (figure 2.5).

GlmS is an amidotransferase that catalyzes the first step of the transformation. It is an essential and dimeric enzyme, whose monomer is composed of two structurally and functionally distinct domains: a glutaminase and an isomerase domain. The structure of GlmS is known in *E. coli* (Protein Data Bank (PDB) entry: 1JXA) (Teplyakov *et al.*, 2001). The N-terminal domain (30 kDa) promotes glu-

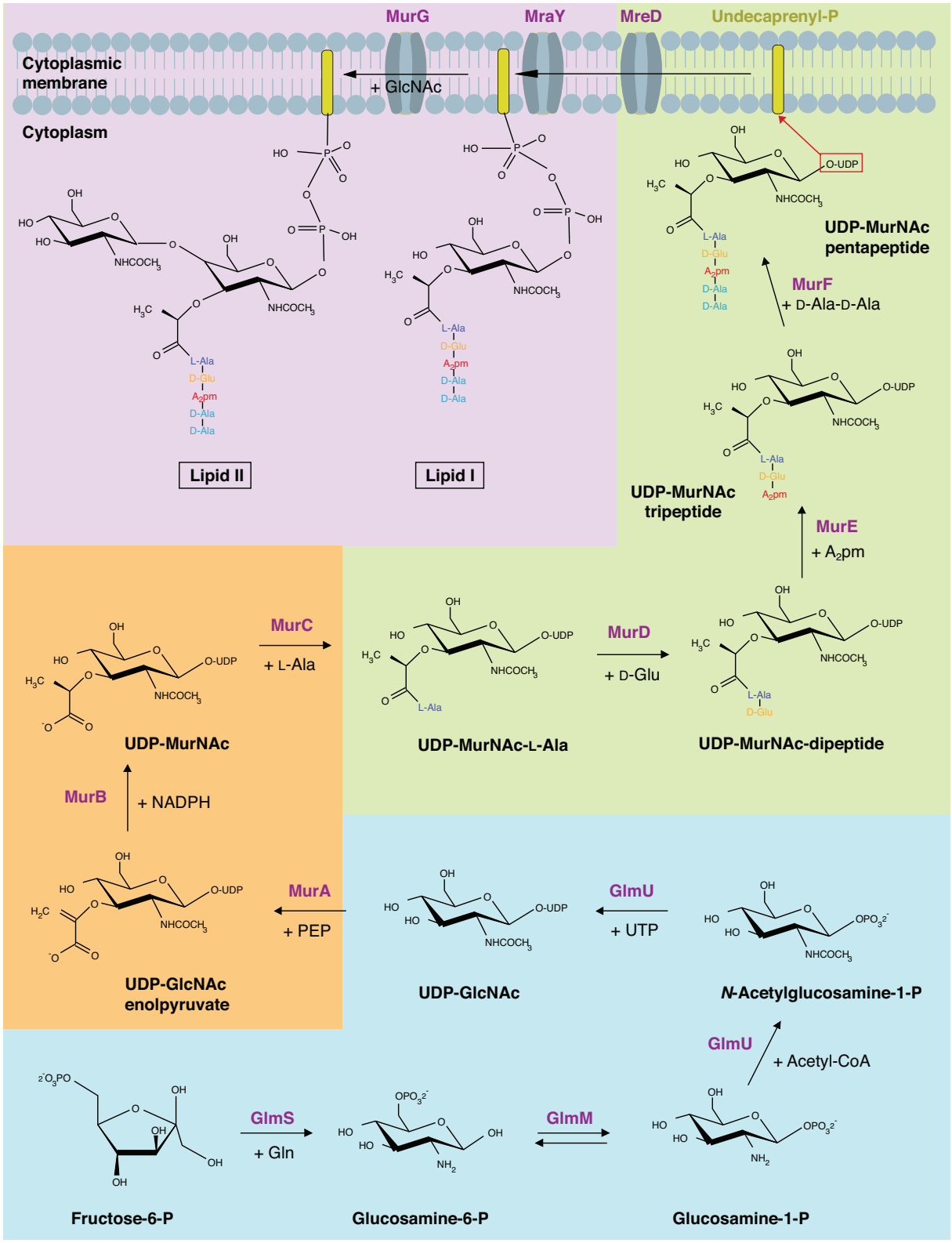


Figure 2.5: Schematic representation of the cytoplasmic steps of peptidoglycan biosynthesis. After the synthesis of precursors UDP-MurNAc (blue and orange backgrounds), the assembly of the pentapeptide stem is ensured by the four Mur ligases: Mur C, Mur D, Mur E and Mur F (green background). This pentapeptide stem is then anchored to the membrane by a linkage with the undecaprenyl-phosphate lipid carrier (in yellow, indicated as undecaprenyl-P), generating Lipid I and Lipid II (mauve background).

tamine hydrolysis into glutamate and ammonia, while the C-terminal domain (40 kDa) binds the nitrogen acceptor and uses ammonia (produced by glutamine hydrolysis) for the conversion of Fructose-6-P into D-glucosamine-6-phosphate (Barreteau *et al.*, 2008).

The second step of the biosynthesis of UDP-GlcNAc is the conversion of glucosamine-6-phosphate into glucosamine-1-phosphate, which is catalyzed by the enzyme Phosphoglucosamine Mutase (PNGM), also called GlmM. GlmM is active only in its phosphorylated form. Structures of two phosphoglucosamine mutases have recently been solved by X-ray, one from *Francisella tularensis* (PDB code: 3I3W, no publication available), and one from *Bacillus anthracis* (PDB code: 3PDK) (Mehra-Chaudhary *et al.*, 2011).

The last step requires the activity of GlmU, a bifunctional enzyme with both acetyltransferase and uridyltransferase activities, as it has been shown by HPLC (Mengin-Lecreulx and van Heijenoort, 1994). First, GlmU catalyzes the acetyl transfer from acetyl coenzyme A (AcCoA) to glucosamine-1-phosphate (Glucosamine-1-P). After the reaction *N*-acetylglucosamine-1-phosphate (Glc-NAc-1-P) is thus released, and the uridyltransfer from UTP to GlcNAc-1-P in the presence of Mg^{2+} finally leads to UDP-GlcNAc (figure 2.5). The structure of GlmU from *E. coli* (PDB entry: 1HV9) (Olsen and Roderick, 2001) shows 2 domains: the acetyltransferase domain at the C-terminal and the uridyltransferase domain at the N-terminal.

Biosynthesis of UDP-MurNAc from UDP-GlcNAc

The first stage towards the creation of the peptidoglycan polymer involves the formation of UDP-MurNAc from UDP-GlcNAc by two enzymes: MurA (formerly known as MurZ), and MurB. The MurA reaction pathway follows an addition-elimination mechanism: a proton is added to the C-3 of Phospho*enol*pyruvate (PEP), whereas the 3'-OH group of UDP-GlcNAc gets deprotonated. After formation of a phospholactoyl-UDP-GlcNAc tetrahedral intermediate (Marquardt *et al.*, 1994), the elimination of phosphate results in the formation of UDP-GlcNAc-*enol*pyruvate. The structures of several forms of MurA have been solved. As an example, the crystal structure of the *E. coli* enzyme MurA complexed with UDP-GlcNAc and fosfomycin has been determined (Skarzynski *et al.*, 1996). The structure consists of two domains with the active site located in between. The two domains have a very similar secondary structure, and the overall protein architecture is similar to that of 5-*enol*pyruvylshikimate-3-phosphate (EPSP) synthase.

Once the UDP-GlcNAc-*enol*pyruvate is produced, it will undergo a reduction catalyzed by MurB that will result in UDP-MurNAc. MurB completes the formation of the muramyl sugar by reducing the *enol*pyruvyl moiety into a lactoyl ether using one equivalent of the reduced DihydroNicotinamide-Adenine Dinucleotide Phosphate (NADPH) (Benson *et al.*, 1996). MurB protein contains a stoichiometric amount of non-covalently bound Flavin Adenine Dinucleotide (FAD) that serves as a redox intermediate. The reduction follows a ping-pong bi-bi mechanism, and is activated by diverse cations. The crystal structures of MurB show that the protein contains three domains: domains 1 and 2 mediate the (FAD) binding, and domain 3 mediates the binding of the reaction substrate. NMR studies on MurB revealed that $NADP^+$ and UDP-GlcNAc-*enol*pyruvate utilize the same binding pocket and that the binding of the latter induces some dynamics in domain 3 (Constantine *et al.*, 1997).

Biosynthesis of UDP-MurNAc peptides

The stepwise assembly of the peptidoglycan peptide stem is ensured by four Mur ligases: MurC, MurD, MurE and MurF. After the consecutive action of MurA and MurB enzymes who generated the UDP-MurNAc precursor, the 4 Mur ligases will generate the stem peptide directly on UDP-MurNAc in presence of a divalent cation, Mg^{2+} or Mn^{2+}. These Mur ligases are essential enzymes that display large similarities:

- They follow the same mechanistic pathway, which consists in the activation of the carbonyl group of the UDP-precursor by ATP, followed by the production of ADP and an acyl phosphate intermediate. Then this intermediate undergoes a nucleophilic attack by the amino group of the added amino acid, leading to the formation of a tetrahedral intermediate.

- They share a series of six invariant protein residues in addition to an ATP-binding consensus sequence, which led to the definition of this class of enzymes as a new family.

- They have similar 3D structures, with 3 domains: an N-terminal region involved in the binding of a UDP-MurNAc precursor, a central domain involved in the recognition of ATP, and a C-terminal domain which recognizes the amino acids that will be added onto the precursor. Furthermore, these enzymes can exist in 'closed' and 'opened' conformations. The 'closed' conformation is favored by the ligand binding.

The MurC ligase adds the first amino acid of the peptide stem, which is a L-Ala in most bacterial species. In some rare cases, MurC can add a L-Ser or a Gly instead of the L-Ala. As an example, the amino acid found at the first position is a Gly in *Mycobacterium leprae* whereas it is a L-Ala in *M. tuberculosis*, although the MurC enzymes of the two species exhibit the same *in vitro* specificity towards these amino-acids. This variation is attributed to differences in growth conditions of the two organisms (Mahapatra *et al.*, 2000). Of note, D-Ala is not a substrate, which means that the reaction is stereospecific.

Then, MurD attaches the γ-D-glutamic acid to the first amino acid. Among the four Mur ligases, MurD differs by its low specificity towards the UDP-precursor. While the three other ligases need an UMP moiety to catalyze the peptide stem assembly, 1-phospho-MurNAc-L-Ala can be a substrate for MurD (Michaud *et al.*, 1987), which means that the reaction is reversible. The second amino acid inserted in the peptide stem is always D-Glu. It may then follow a post-modification through amidation into a D-Gln, as it is the case in most Gram-positive bacteria and mycobacteria (Table 2.2).

The third amino-acid is assembled by the MurE enzyme, and is generally either *meso*-A_2pm in most Gram-negative bacteria, or L-lysine in most Gram-positive bacteria (Barreteau *et al.*, 2008). *meso*-A_2pm is the 2,6-diaminopimelic acid and is also called *meso*-DAP. The most significant variations among organisms are found in this third position, as shown in Table 2.2.

Finally, MurF adds a dipeptide D-Ala,D-Ala in most cases, to yield the final molecule UDP-MurNAc-pentapeptide. In strains that are resistant to vancomycin, D-Lactate (D-Lac) or D-Ser can replace the D-Ala in position 5, yielding to a lower antibiotic affinity (Healy *et al.*, 2000).

The amino acids conventionally present at each position, and identified alterations are detailed in Table 2.2.

Position	Residue encountered	Examples
1	L-Ala	Most species
	Gly	*M. leprae*, Brevibacterium imperiale
	L-Ser	Butyribacterium rettgeri
2	D-isoglutamate	Most Gram-negative species
	D-isoglutamine	Most Gram-positive species, Mycobacteria
	threo-3-hydroxyglutamate	*Mycobacterium lacticum*
3	*meso*-A_2pm	Most Gram-negative species, Bacilli, Mycobacteria
	L-Lys	Most Gram-positive species
	L-Orn	*Spirochetes, Thermus thermophilus*
	L-Lys / L-Orn	*Bifidobacterium globosum*
	L-Lys / D-Lys	*Thermotoga maritima*
	LL-A_2pm	*Streptomyces albus, Propionibacterium petersonii*
	meso-Lanthionine	*Fusobacterium nucleatum*
	L-2,4-Diaminobutyrate	*Corynebacterium aquaticum*
	L-Homoserine	*Corynebacterium poinsettiae*
	L-Ala	*Erysipelothrix rhusiopathiae*
	L-Glu	*Arthrobacter J. 39*
	Amidated *meso*-A_2pm	*Bacillus subtilis*
	2,6-diamino-3-hydroxypimelate	*Ampuraliella regularis*
	L-5-hydroxylysine	*Streptococcus pyogenes*
	N-acetyl-L,2,4-diaminobutyrate	*Corynebacterium insidiosum*
4	D-Ala	All bacteria
5	D-Ala	Most bacteria
	D-Ser	*Enterococcus gallinarum*
	D-Lac	*Lactobacillus casei, Enterococci resistant to vancomycin*

Table 2.2: Amino acid variations in the peptide stem of peptidoglycan from different species. The variations in peptide stem composition can be divided into two classes: variations occurring during the peptidoglycan assembly due to the specificity of Mur ligases, and modifications posterior to synthesis due to growth conditions (Vollmer *et al.*, 2008) or to antibiotics administration (Mainardi *et al.*, 2008). Additional modifications of the peptide stem occurring after the action of Mur ligases, often at the level of Lipid II, include amidation, hydroxylation, acetylation, and/or additional amino acid attachment. This table is taken from (Vollmer *et al.*, 2008).

2.2.2 Formation of Lipid II and transport to the perisplasm

Biosynthesis of Lipid I

Undecaprenyl phosphate, also known as bactoprenol or C_{55}-P as it contains 55 carbons, is the classical carrier lipid encountered in the bacterial world, and is the key lipid involved in the biosynthesis of peptidoglycan. It is essential for the assembly of the UDP-GlcNAc and UDP-MurNAc-pentapeptide precursors and for the transport of hydrophilic GlcNAc-MurNAc-peptides across the hydrophobic environment of the cytoplasmic membrane towards the periplasm. The transfer of the MurNAc-pentapeptide moiety phosphorylated at position 1 from the soluble UDP nucleotide precursor onto the C_{55}-P carrier lipid was demonstrated for the first time in 1965, using membrane preparations from *S. aureus* and *Mycrococcus luteus* (Anderson *et al.*, 1965; Struve and Neuhaus, 1965). This reaction consists in the translocation of the soluble phosphorylated MurNAc-pentapeptide onto the C_{55}-P carrier lipid and is catalyzed by the translocase MraY (van Heijenoort, 2001b). MraY was first difficult to purify and characterize compared to the highly soluble Mur enzymes. It is only in 2004, 40 years after the identification of the *mraY* gene, that MraY could be characterized (Bouhss *et al.*, 2004), opening the way to structural analysis of this essential membrane protein. Yet, no structure of MraY has been solved to date.

Biosynthesis of Lipid II

The translocase II (MurG) catalyses the second membrane-associated step in the peptidoglycan synthesis. This step involves the transfer of a GlcNAc molecule from UDP-GlcNAc to Lipid I thanks to the MurG enzyme, forming the lipid-anchored disaccharide-pentapeptide monomer subunit, also called Lipid II. Lipid II is the unique membrane-associated precursor essential for cell wall growth (Bouhss *et al.*, 2008; Mohammadi *et al.*, 2011). MurG is a GlycosylTransferase (GT) belonging to the GT-B superfamily. The X-ray structure of MurG in *E. coli* (PDB entry 1F0K) (Ha *et al.*, 2000) reveals that the free enzyme contains two domains separated by a deep cleft. Both domains exhibit high structural similarities despite minimal sequence homology. According to this structure and to sequence homology with other Mur enzymes, the C-terminal domain of *E. coli* MurG would be the UDP-GlcNAc binding site (Ha *et al.*, 2000). In 2003, the structure of *E. coli* MurG in complex with UDP-GlcNAc was published (Hu *et al.*, 2003) and showed that the binding of the Lipid I substrate was likely to take place in the cleft between the two domains.

In addition, recent results on *Caulobacter crescentus* suggest that the synthesis of the membrane-bound peptidoglycan precursors and their transfer through the inner membrane may involve additional proteins and may require the assembly of a complex machinery (White *et al.*, 2010). This study shows in particular that the integral membrane protein MreD is essential for lateral peptidoglycan synthesis, interacts with the precursor synthesizing enzymes MurG and MraY, and additionally, determines the actin-like protein MreB localization. The supramolecular organization of these proteins with the membrane proteins RodA and RodZ, and the periplasmic MreC protein would be required for propagation of a uniform cell shape and would facilitate the transport of Lipid II towards the periplasm.

Flipping of Lipid II across the inner membrane

How can Lipid II flip from the intracellular to the extracellular face of the membrane? If the en-

zymatic cytoplasmic steps are relatively well known, the membrane step involving the Lipid II transfer from the cytoplasm to the perisplasm is not well understood. This process might involve enzymes called "flippases". Recent studies have shown that membrane proteins, such FtsW (Mohammadi *et al.*, 2011) and/or RodA (Höltje, 1998; Ikeda *et al.*, 1989) which are part of the SEDS family (Shape, Elongation, Division and Sporulation family), are required for the transport of lipid-linked cell wall precursors across biogenic membranes. FtsW and RodA show a high degree of similarity and are predicted to have 10 trans-membrane-spanning regions. An interaction with PBP3 and PBP2, respectively, which are both specific murein synthases, has been shown (Matsuhashi *et al.*, 1990), revealing that these two membrane proteins may be part of the machinery of the membrane step of peptidoglycan biosynthesis (Typas *et al.*, 2012).

2.3 Polymerization of nascent peptidoglycan in the periplasm

Once on the periplasmic side of the membrane, Lipid II undergoes two reactions: the transglycosylation, a polymerization of the glycan chains catalyzed by glycosyltransferases (GTs), and the transpeptidation, a reaction that cross-links the peptide stems with the glycan chains and that is catalyzed by D,D-TransPeptidases (DDts) or the newly-discovered L,D-TransPeptidases (LDts) (figure 2.6).

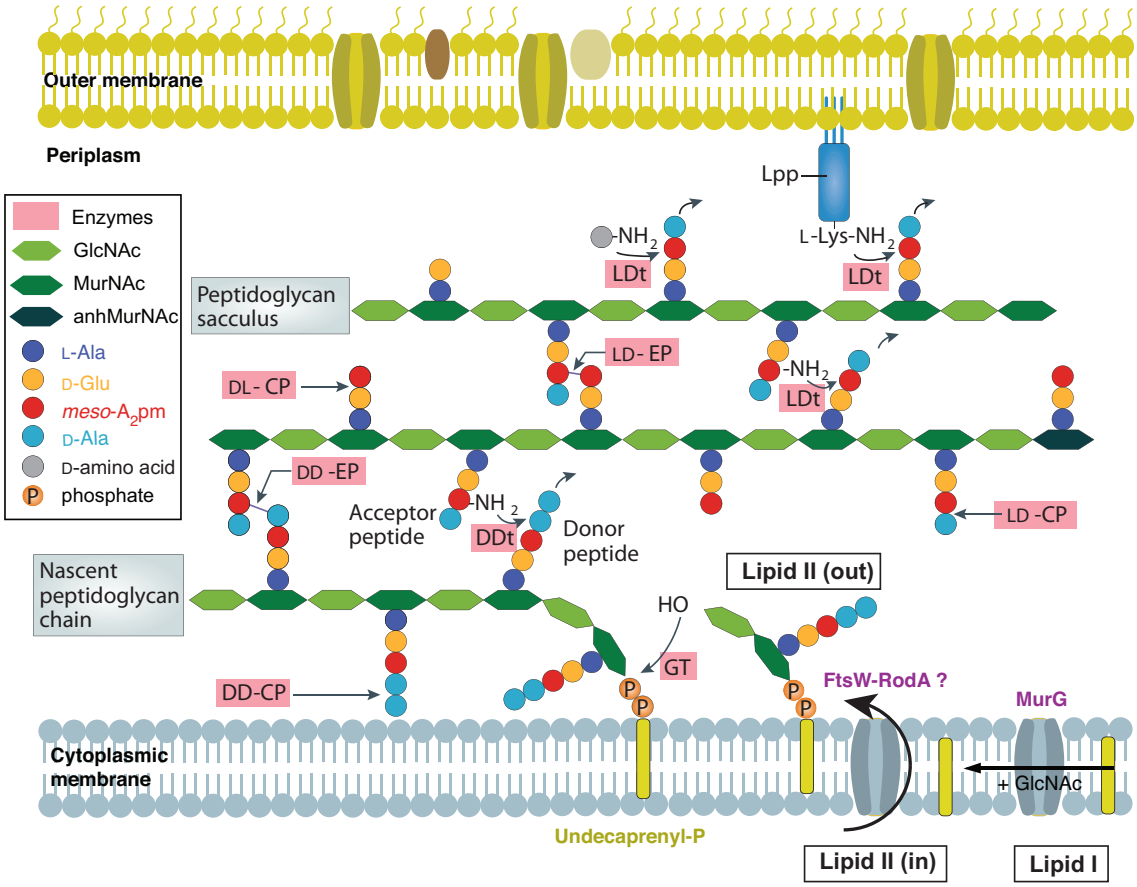


Figure 2.6: Peptidoglycan synthesis and hydrolysis occurring in the *E. coli* periplasm. Figure adapted from (Typas *et al.*, 2012). Glycosyltransferases (GTs) catalyze the formation of a nascent peptidoglycan chain from the Lipid II precursor. This reaction is followed by the transpeptidation reaction, with D,D-transpeptidases (DDts) that catalyze the peptide bond formation between D-Ala^4 of the donor stem and the *meso*-A$_2$pm^3 of the acceptor stem, or with L,D-transpeptidases (LDts) that catalyze the peptide bond formation between *meso*-A$_2$pm^3 of the donor stem and *meso*-A$_2$pm^3 of the acceptor stem. LDts are also responsible for the attachment of the major lipoprotein of *E. coli* (Lpp), which is anchored in the outer membrane. Cross-links are cleaved by D,D and L,D-endopeptidases (EPs), and peptides are trimmed by carboxypeptidases (CPs). D,D-CPs trim the bond between the D-Ala^4 and D-Ala^5 terminus of the pentapeptide stem of peptidoglycan, L,D-CPs trim the bond between *meso*-A$_2$pm^3 and D-Ala^4 of the tetrapeptide, and D,L-CPs trim the bond between the D-Ala^2 and the *meso*-A$_2$pm^3 of the tripeptide. The role of these peptide stem hydrolysis in the periplasm performed by CPs is to regulate the amount of peptidoglycan.

2.3.1 Polymerization via the PBPs and D,D-transpeptidases

PBPs were initially identified because of their ability to covalently bind penicillin (Tipper and Strominger, 1965), which gave them their actual name. PBPs are anchored to the cytoplasmic membrane, and domains with catalytic activity are located on the periplasmic extracellular side. These enzymes belong to the ASPRE (Active-Site Serine Penicillin Recognizing Enzymes) super family, containing transpeptidases, carboxypeptidases and β-lactamases. Members of this family are characterized by common SXN, SXXK and K(T/S)G structural motifs (Goffin and Ghuysen, 1998).

Thanks to the achievement of the sequencing of many bacterial genomes, the number of PBPs per organism has been determined (Sauvage *et al.*, 2008). Each bacterial species possess a variable number of PBPs, generally at least 3 and up to more than 8 can be present. As an example, *E. coli* comprises three bifunctional synthases (PBP1a, PBP1b, and PBP1c), one glycosyltransferase (MgtA) and two transpeptidases (PBP2 and PBP3) involved in cell elongation and in cell division, respectively (Nanninga, 1998). Whereas the inactivation of either PBP1a or PBP1b is tolerated, the simultaneous absence of both enzymes results in cell lysis (Yousif *et al.*, 1985).

2.3.1.1 Structural classification of Penicillin Binding Proteins

PBPs can harbor both transglycosylation and transpeptidation activities (Sauvage *et al.*, 2008), or only one of them. They have thus been classified into two major categories: the High Molecular Mass (HMM) PBPs and the Low Molecular Mass (LMM) PBPs. HMM PBPs are further divided into two classes : class A enzymes, which catalyze both GT and TransPeptidase (TP) activities, and class B enzymes, which are limited to the transpeptidation catalysis. LMM PBPs can exert carboxypeptidase and/or endopeptidase activities and thus permit to regulate the level of peptidoglycan cross-linking by removing the terminal D-Ala from the muramyl peptide. Yet, in contrast to HMM PBPs, they are not essential for normal cell growth and viability (Sauvage *et al.*, 2008).

HMM PBPs are composed of a short N-terminal cytoplasmic region, a trans-membrane domain (TM domain) which is anchored in the inner membrane, and a C-terminal catalytic domain which harbors the transpeptidase activity (TP domain) (figure 2.7). In the case of class A PBPs, the N-terminal membrane-anchored domain contains glycosyltransferase activity (GT domain), whereas the function of the equivalent domain in class B PBPs is not known (non-Penicillin Binding (n-PB) domain).

Class A PBPs

The analysis of the sequences of 63 PBPs showed that among the 29 PBPs of class A, 6 or 7 subfamilies could be distinguished (Goffin and Ghuysen, 1998; Sauvage *et al.*, 2008): three clusters for Gram-negative bacteria, represented by PBP1a-c from *E. coli*, three clusters for Gram-positive bacteria, whose prototypes are *B. subtilis* PBP2c, *S. pneumoniae* PBP1a and PBP1b, and one category for mycobacteria represented by *M. tuberculosis* PBP1a. This sequence analysis revealed that there are significative variations in the TP domains depending on the bacterial species, whereas the GT domain is well-conserved. This conservation might be due to the composition of peptidoglycan glycan chains, which is highly identical between species, and thus does not need any adaptation of the polymerization enzymes. In contrast, the TP substrates (the pentapeptides), differ significantly between species and are dependent on growth conditions, as discussed in section 2.1.1 and shown in Table 2.2.

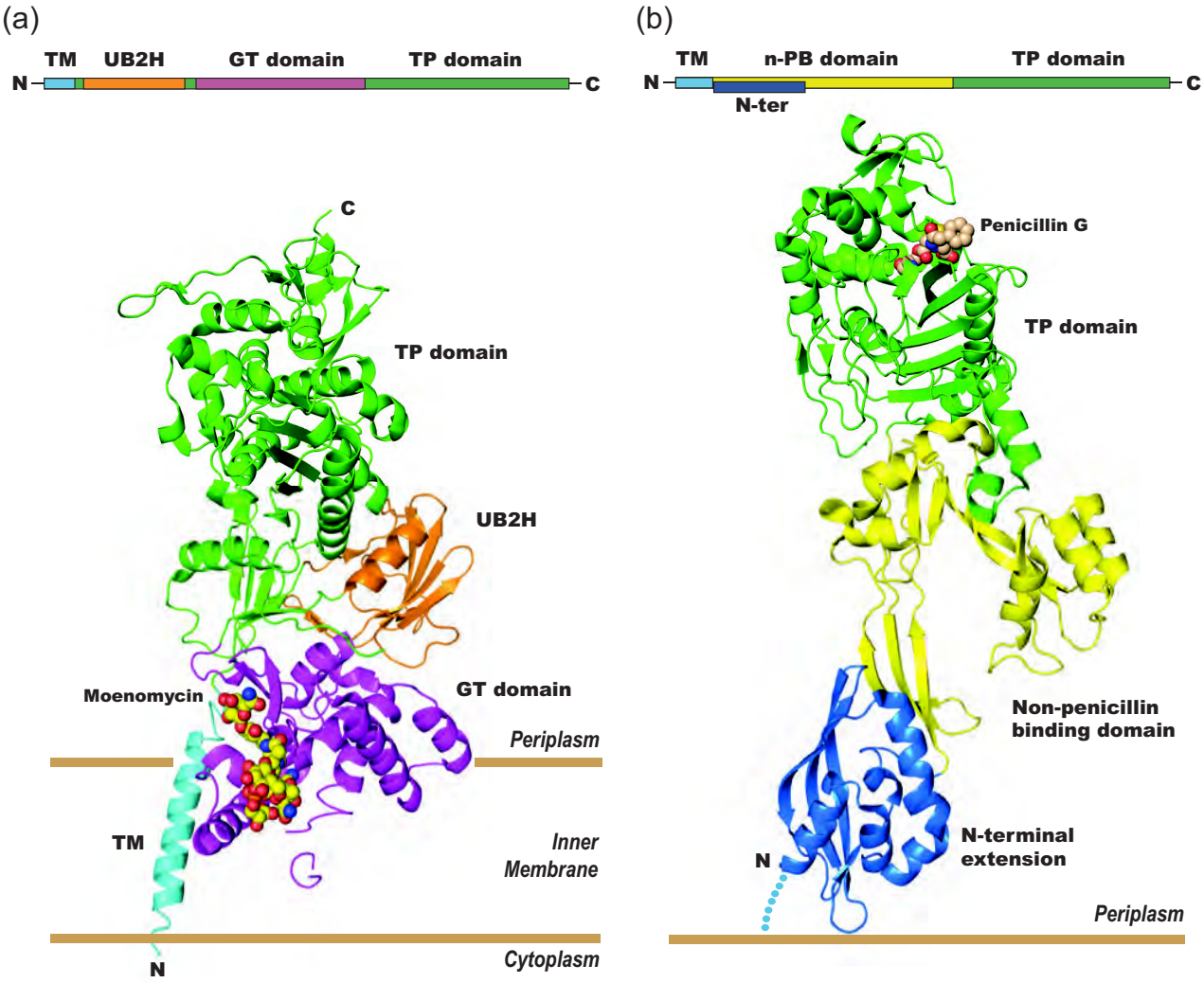


Figure 2.7: Crystal structures of (a) full-length *E. coli* PBP1b from class A acylated by moemomycin (PDB entry: 3VMA) (Sung *et al.*, 2009) and **(b) *S. aureus* PBP2a from class B** acylated by penicillin-G (PDB entry: 1MWT) (Lim and Strynadka, 2002). The TM and TP domains are shared between the two PBP classes and are colored in cyan and green, respectively (TM domain is not visible on PBP2a structure). GT and UB2H domains of PBP1b are shown in magenta and orange. For PBP2a the N-terminal nPB domain is shown in yellow, with the N-terminal extension colored in blue. The inhibitors, moenomycin in (a) and penicillin-G in (b) are represented as spheres.

On the structural level, class A PBPs have been more difficult to study compared to class B enzymes, because of the hydrophobic character of GT domains. To date, only two isolated GT domain structures (Barrett *et al.*, 2007) and two full-length class A PBPs structures have been solved by crystallography. The full-length structures both include the transmembrane domain, as shown in figure 2.7a, which represents the *E. coli* PBP1b in complex with moenomycin (Sung *et al.*, 2009). The other full-length structure of class A PBP deals with PBP2 of *S. aureus* (Lovering *et al.*, 2007). Structures were both solved on the apo- and moenomycin-acylated forms. In addition, a model for the structure of PBP1b from *S. pneumoniae* was recently published using a combination of SAXS and crystallographic data (Macheboeuf *et al.*, 2011). If these models and structures can provide important knowledge regarding antibiotic development, it does not solve the question about the substrate recognition, as the interactions and flexibility are not easily understood from the X-ray structures in the absence of peptidoglycan substrates.

Class B PBPs

All the class B PBPs share the presence of a catalytic domain TP responsible for the cross-linking of peptidoglycan peptide stems, and an N-terminal domain with an unknown function. PBP2x from *S. pneumoniae* was the first structure of HMM PBP of class B, solved in 1996 with a resolution of 3.5 Å (Pares *et al.*, 1996). This structure revealed that the N-terminal domain (equivalent of domains in blue and yellow in figure 2.7b) presented an unknown shape similar to a "pair of sugar tongs" with a hole of $\sim$ 10 Å diameter, without equivalent at that time in the structural data banks. The length of this domain can vary from 60 residues in *Deinococcus radiodurans* to 400 residues on *Clostridium tetani*. In 2000, the structure of PBP2x from *S. pneumoniae* was solved again with a better resolution for both apo- and acylated forms of the enzyme (Gordon *et al.*, 2000). It revealed that the active site of the acylenzyme is more disordered than that of the apoenzyme. Some residues to the left of the active-site cleft are poorly resolved and some others, towards the base of the active site, appear to have moved. The loop comprising residues 555 to 565 is poorly ordered in both structures, indicating a degree of flexibility independent of antibiotic binding, which suggested that the muropeptide, natural substrate of PBP2x, could possibly interact with this loop (Gordon *et al.*, 2000).

PBPs of low molecular mass

LMM PBPs are membrane-associated and monofunctional enzymes that catalyze the D,D-carboxypeptidase activity. They play important roles in the maintenance of cell shape and in cellular growth and division processes. LMM PBPs contain a TP domain prior to a cleavable signal peptide and a transmembrane domain enabling them to anchor in the membrane through a C-terminal amphiphilic α-helix. Some high resolution structures of LMM PBPs have been solved, including the ones of *E. coli* PBP5 (PDB entry: 1NZ0) (Nicholas *et al.*, 2003), *S. aureus* PBP4 (PDB entry: 1TVF, but no publication available) and *S. pneumoniae* PBP3 (PDB entry: 1XP4) (Morlot *et al.*, 2005). PBP5 from *E. coli* and PBP3 from *S. pneumoniae* have been both suggested to be vertically positioned on the bacterial membrane. This orientation would be possible thanks to the β-rich domain separating the catalytic region from the amphiphatic region, also called domain II in PBP3. It would serve as a pedestal that would bring the active site close to the peptidoglycan layer. The peptidic side-chains of peptidoglycan could therefore interact with LMM PBPs and may extend in the direction of the cytoplasmic membrane (Morlot *et al.*, 2005).

2.3.1.2 Catalytic activity of Penicillin Binding Proteins

Transglycosylation

On the contrary to the transpeptidation, for which there is an abundant literature, the transglycosylation step has been studied to a far less extent. Two factors can account for this: first the difficulty to handle GT enzymes as they are membrane proteins; second, the limitation in appropriate substrates availability as Lipid II is present in very small quantities in bacterial cells and cannot be isolated in large quantities from natural sources. Indeed, the UDP-MurNAc-pentapeptide/ Lipid II ratio was estimated at ca. 100:1 (Kohlrausch *et al.*, 1989) or 140:1 (Mengin-Lecreulx *et al.*, 1991). Considering that there are ca. 10^5 molecules of UDP-MurNAc-pentapeptide per cell, there would thus be only 1,000 to 2,000 molecules of Lipids II (van Heijenoort *et al.*, 1992). Such extremely low natural quantities explain why the isolation and purification of Lipid II directly from cells is not an easy matter. This limitation is slowly evolving with the in-vitro preparation of a ^{14}C-[GlcNAc]-labeled Lipid II (Bertsche *et al.*, 2005).

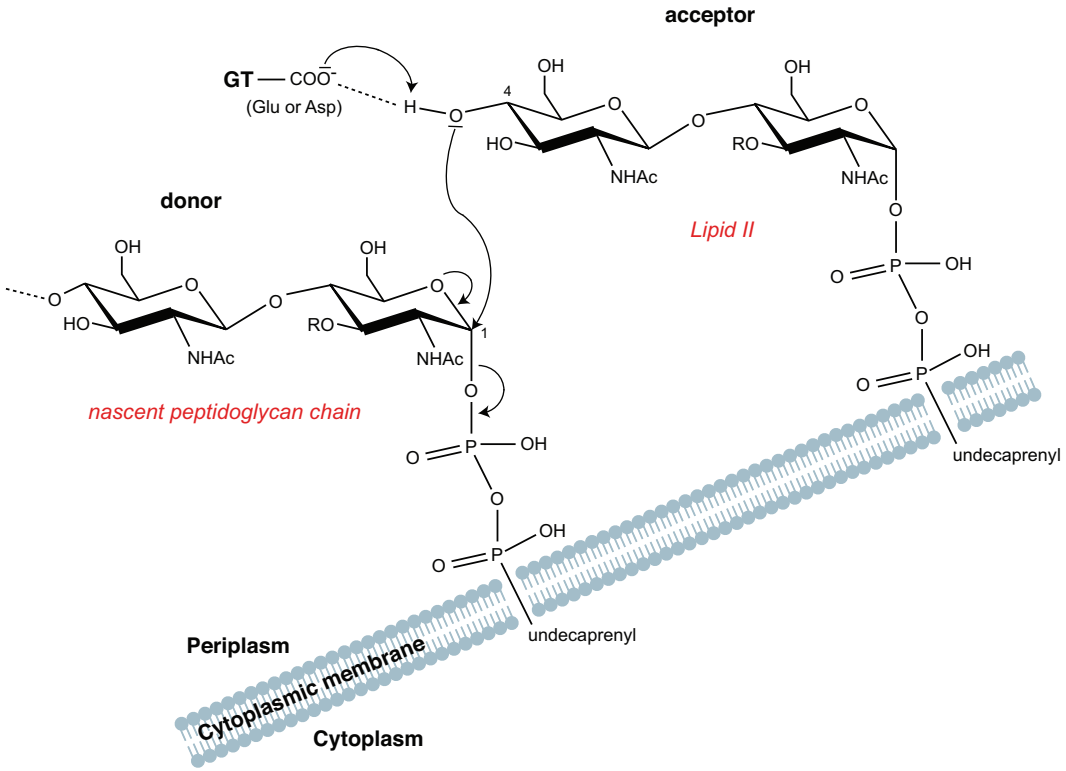


Figure 2.8: Representation of the extension of glycan chains by GlycosylTransferases (GTs). Elongation occurs with the nascent peptidoglycan chain as the donor and Lipid II as the acceptor.

To date, two glycosylation mechanisms have been described in the literature (van Heijenoort, 2001a). The first studies have shown that the growing glycan chain, anchored to the membrane through the undecaprenyl phosphate on the last MurNAc, would be the donor and would be transferred to the acceptor Lipid II molecule, onto the hydroxyl group at position 4 on the GlcNAc entity, as depicted in figure 2.8. This mechanism was first established on two Gram-positive bacteria: *Bacillus licheniformis* (Ward and Perkins, 1973) and *Micrococcus luteus* (Weston and Perkins, 1977). An alternative mechanism arose many years later, in which the Lipid II was considered as the donor substrate, and the disaccharide-pentapeptide would be transferred towards the acceptor nascent peptidoglycan (van Heijenoort, 2001a).

However, all the recent studies are in favor of the primarily formulated hypothesis (Welzel, 2005; Lovering *et al.*, 2008), especially since structural data were obtained on GlycosylTransferases.

The first structure of peptidoglycan GT (PGT) to be reported was the PGT domain of the class A PBP1a from *Aquifex aeolicus* (Yuan *et al.*, 2007). The 2.1 Å X-ray structure revealed an almost completely α-helicoidal fold, which drastically differed from all other glycosyltransferase structures reported at that time. Yet, some resemblance to the λ lysozyme were detected, an enzyme that degrades the carbohydrate chains of peptidoglycan. This structure confirmed the first model for glycan chain polymerization, and showed in addition that PGTs are processive, which means that they catalyze multiple rounds of coupling without releasing the elongating product (Barrett *et al.*, 2007; Yuan *et al.*, 2007).

Furthermore, the second structure published soon after the first one for PBP2 from *S. aureus* (Lovering *et al.*, 2007) also confirmed the transglycosylation model with the Lipid II acting as the acceptor. On the basis in particular of the complex structure with moemomycin, the authors proposed that the antibiotic molecule in the binding site of the transglycosylase structurally mimics Lipid IV, a Lipid II dimer. This model would place the C1-carbon of the MurNAc saccharide from the growing nascent peptidoglycan chain at the center of the active site, confirming its role as the donor. The catalytic Glu^{114} would then deprotonate the hydroxyl group in position 4 of the GlcNAc of the Lipid II acceptor, triggering a concomitant attack on the C1-carbon of the growing nascent peptidoglycan chain to form a β-1,4-linked product (figure 2.8) (Lovering *et al.*, 2008).

Transpeptidation and carboxypeptidation

The carboxypeptidation and transpeptidation reactions catalyzed by PBPs involve the peptide stems of the peptidoglycan only and follow a three-step mechanism (figure 2.9).

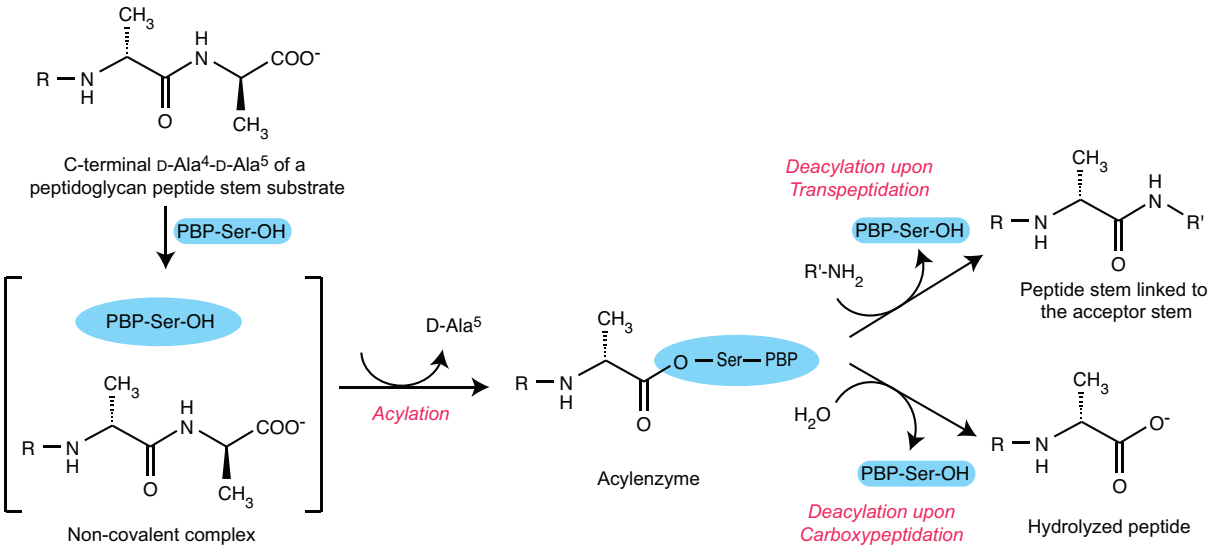


Figure 2.9: Schematic representation of the transpeptidation and carboxypeptidation reactions performed by PBPs.

First there is a reversible formation of a non covalent Michaelis complex between the enzyme and the donor strand of peptidoglycan, the [L-Ala^1-γ-D-Glu^2-*meso*-A_2pm^3 (or L-Lys^3)- D-Ala^4- D-Ala^5] peptide stem. Then, the enzyme acylation occurs through the nucleophilic attack of the carbonyl carbon of the D-Ala^4 by the PBP active serine (Ghuysen *et al.*, 1973; Rasmussen and Strominger, 1978). The C-

terminal D-Ala^5 is concomitantly released in this process. The name of the enzyme, D,D-transpeptidase originated in the cleavage of this D-Ala-D-Ala peptide bond. The final step in both reactions results in the deacylation of the enzyme. It consists either in hydrolysis, with release of the shortened peptide (carboxypeptidation), or in cross-link formation with the amino acid in position 3 of a second peptidoglycan stem peptide called the acceptor stem (transpeptidation) (figure 2.9).

A recent computational study conducted on *S. pneumoniae* PBP1b (Shi *et al.*, 2011) suggests that the non-covalent Michaelis-Menten complex is stabilized by an hydrogen-bond network involving the carbonyl of D-Ala^4, Thr^{454} and the catalytic residues of the STTK motif (Ser^{460} and Lys^{463}). The same authors proposed that Lys^{463} helps in the deprotonation of the catalytic Ser^{460} to yield a tetrahedral intermediate where the latter residue is covalently bound to the D-Ala^4 former carbonyl group. The protonated lysine then becomes the catalytic acid enabling the collapse of the acylenzyme through a water molecule hydrogen-bonded to the leaving D-Ala^5. During the third and last step of the transpeptidation reaction, the nucleophilic attack of the acylenzyme by the amino group of the donor stem, an unfavorable new *cis*-peptide bond is formed, which is energetically disfavored. During this third step, Lys^{463} favors the deprotonation of the nucleophilic amine and Thr^{454} stabilizes the oxyanion. The isomerization of the *cis*-peptide bond into a *trans* configuration while bound to the enzyme is the rate limiting step of the mechanism as it requires a conformational transition for the enzyme. This isomerization may trigger the product release to complete enzymatic turnover. The role of the two additional consensus motifs SWN and KTG is less clear. They may contribute to the width and linearity of the cleft formed at the active site to accommodate the two reacting peptide stems.

2.3.2 L,D-transpeptidation, an alternative transpeptidation pathway

2.3.2.1 Activation of a new pathway for peptidoglycan polymerization

Discovery of peptidoglycan 3→3 cross-link

The first "unusual" pathway for peptidoglycan cross-linking in non-coryneform bacteria was detected in *E. coli* in 1988 (Blasco *et al.*, 1988). In this study, the authors explored the chemical composition of the peptidoglycan through HPLC analysis in the absence and in the presence of cephalosporin or penicillin antibiotics. They evidenced significant alterations, with the presence of *meso*-A$_2$pm^3 → *meso*-A$_2$pm^3 (3→3) cross-links instead of the classical D-Ala^4 → *meso*-A$_2$pm^3 (4→3) cross-links generated by the D,D-transpeptidase activity of PBPs. If this alternative cross-linking was observed in minor proportions in exponential (2%) and stationary (4%) growth phases, these ratios could be significantly affected by the presence of mecillinam (a penicillin antibiotic) in the growth medium (6.9% of 3→3 cross-links in the stationary phase).

In 2000, the 3→3 cross-link was also detected in *E. faecium*, in an ampicillin-resistant strain called M512 (Mainardi *et al.*, 2000). This resistant strain was obtained from the *E. faecium* D344S parental strain, that derives from the WT-strain by a spontaneous deletion of the low affinity PBP, PBP5. It is known that among the *E. faecium* D,D-transpeptidases, low affinity PBP5 from class B is responsible for low-levels of resistance to β-lactams. Deletion of the *pbp5* gene thus provides the D344S parental strain with a high susceptibility to ampicillin. The D344S strain was grown on increasing concentrations of ampicillin. After five selection steps, no inhibitory effect was observed on the bacterial growth of the

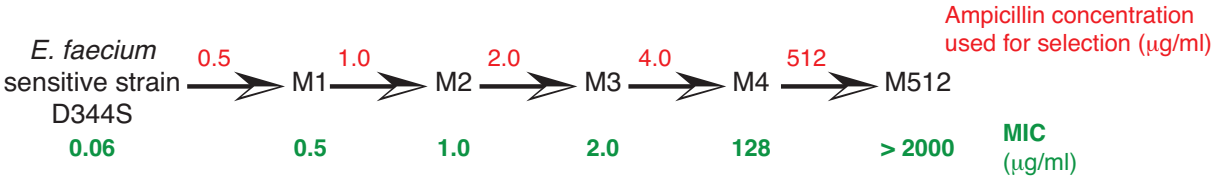


Figure 2.10: Minimum Inhibitory Concentration (MIC) values of mutant strains from *E. faecium*. After five steps, the resistance to ampicillin evolved from 0.06 to 2,000 μg/ml, generating a highly ampicillin resistant strain, called M512. Data taken from (Mainardi *et al.*, 2000, 2002).

M512 mutant strain after overnight incubation up to the highest ampicillin concentration tested, 2,000 μg/ml, a Minimum Inhibitory Concentration (MIC) that is approximately 30,000 times that of the parental D344S strain (figure 2.10).

It was shown that L-Lys^3 → D-Asx-Lys^3 cross-link was present in 100 % of the muropeptide of the ampicillin-resistant strain M512, instead of the D-Ala^4 → D-β-Asx-Lys^3 observed in the parental D433S strain (Mainardi *et al.*, 2000). For the first time the D,D-transpeptidase activity was shown to be unessential, and was thus by-passed by the LDt activity to perform peptidoglycan polymerization. This by-pass does not require an over-expression of these enzymes, as the L,D-transpeptidase activity was similar in

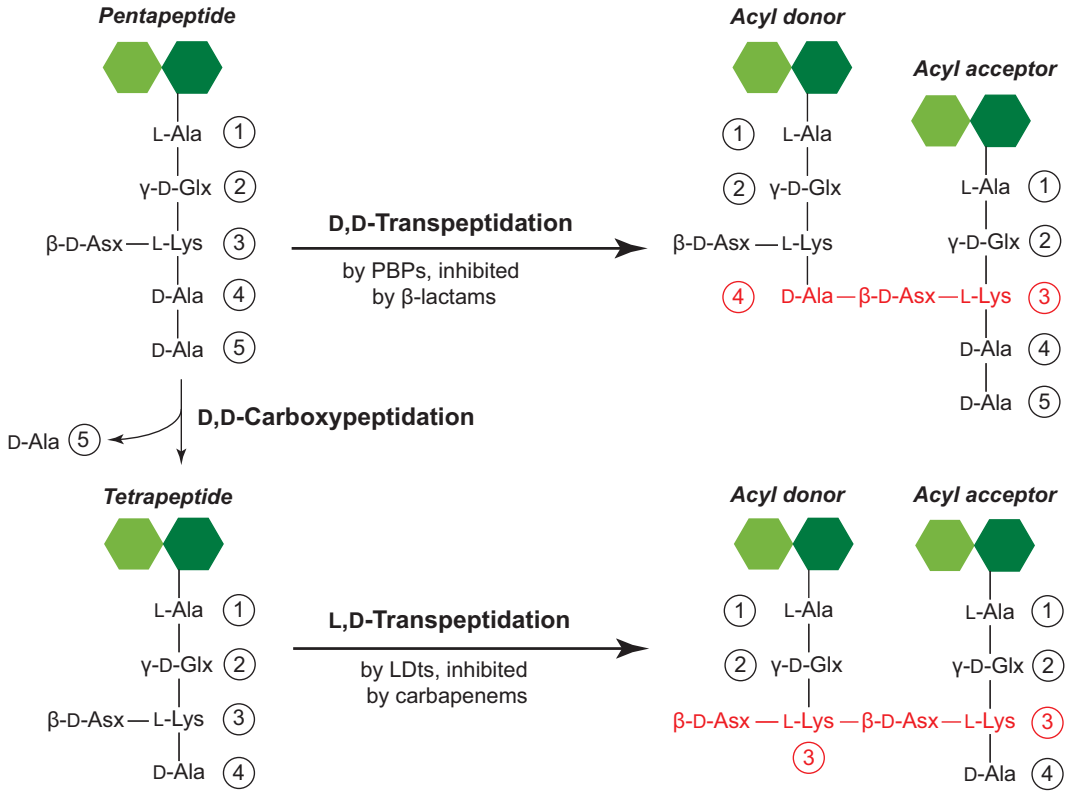


Figure 2.11: Schematic representation of the cross-links generated by transpeptidation in *E. faecium*. D,D-transpeptidation is catalyzed by PBPs which link the D-Ala^4 to L-Lys^3 (4→3 cross-link) through the bridging β-D-Asx (the Asx backbone amide is the acceptor site for the D-Ala^4 backbone carbonyl of the donor stem. Asx represents Asn or Asp residues). In β-lactam resistant strains, D,D-carboxypeptidases activate the L,D-transpeptidation pathway by overproduction of disaccharide tetrapeptide precursors through the cleavage of D-Ala^5. LDts then link the L-Lys^3 from the donor stem to the L-Lys^3 of the acceptor stem (3→3 cross-link) through the bridging β-D-Asx (the Asx backbone amide is the acceptor site for the L-Lys^3 backbone carbonyl of the donor stem). Hexagons represent the GlcNAc (light green) and the MurNAc (dark green) molecules.

both D344S and M512 strains. In fact, it is the production of β-lactam insensitive D,D-carboxypeptidases that activates the L,D-transpeptidation pathway through overproduction of the tetrapeptide precursor, the natural LDt substrate (Mainardi *et al.*, 2002), as emphasized in figure 2.11. At the same time, the D,D-carboxypeptidase activity inhibits the D,D-transpeptidation pathway because tetrapeptides (as opposed to pentapeptides) are non-substrates for PBPs. The availability of tetrapeptide stems is therefore one of the limiting factors to activate this resistance pathway.

Identification of L,D-transpeptidases

In 2005, the enzyme responsible for the L,D-transpeptidation pathway in *E. faecium* was identified by reverse genetics on the M512 strain (Mainardi *et al.*, 2005). Using a radioactive assay on tetrapeptide peptidoglycan precursors, a soluble enzymatically active fraction was partially purified in four chromatographic steps and analyzed on Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE). The Open Reading frame (ORF) coding for a 48 kDa protein responsible for this activity was subsequently identified after N-terminal sequencing and a Blast search through the *E. faecium* genome. The corresponding recombinant protein was produced in *E. coli* and was shown to catalyze the L,D-transpeptidation and not to be inhibited by ampicillin. These results confirmed the successful identification of the *E. faecium* L,D-transpeptidase (Ldt_{fm}) gene. While serine was identified in PBPs as the catalytic residue, site directed mutagenesis of Cys^{442} in Ldt_{fm} into Ala yielded an inactive protein, thus suggesting this residue as the L,D-transpeptidation catalytic site.

After the discovery of the novel 3→3 pathway, followed by the identification of Ldt_{fm}, genomes were screened in order to evaluate the distribution of cysteine-active L,D-transpeptidases among eubacteria. More than 350 protein sequences were identified which contained cysteine C-terminal sequences similar to that of Ldt_{fm} catalytic domain, designated as ErfK/YbiS/YnhG domain (Pfam PF03734) (Mainardi *et al.*, 2005). Some of the corresponding recombinant proteins were since then produced and functionally characterized in order to challenge the idea according which these cysteine-active catalytic domains were specifically catalyzing the formation of 3→3 cross-links between peptidoglycan tetrapeptide precursors.

Among them, two additional LDts were characterized in *E. faecalis* (Ldt_{fs}) and in *Bacillus subtilis* (Ldt_{Bs}) in 2007 (Magnet *et al.*, 2007a). Each of the three LDts, Ldt_{fm}, Ldt_{fs} and Ldt_{Bs}, proved to be specific for its natural disaccharide tetrapeptide chemotype *in vitro*, respectively L-Lys^3- D-iAsn, L-Lys^3- L-Ala-L-Ala and *meso*-A_2pm^3. While Ldt_{Bs} and Ldt_{fm} accommodated variations in the chemical structure of the acceptor stem peptide at position 3, Ldt_{fs} was extremely specific and adopted additional L,D-carboxypeptidase and D,D-endopeptidase activity that were not present in the other two enzymes.

In *Corynebacterium jeikeium* (*C. jeikeium*), an emerging nosocomial pathogen, two proteins with a LDt-like catalytic domain have been identified (Lavollay *et al.*, 2009), namely Ldt_{Cjk1} and Ldt_{Cjk2}. Ldt_{Cjk1} catalyzes the formation of 3→3 cross-links between the *meso*-A_2pm^3 residues of two distinct peptide stems as expected, while Ldt_{Cjk2} only catalyzes the exchange of the D-Ala^4 in the disaccharide tetrapeptide. Of note, Ldt_{Cjk1} is responsible for 38 % of the cross-links in *C. jeikeium* (the other 4→3 cross-links arising from the low affinity PBP2c activity), a much higher percentage than in the formerly described bacterial species.

More recently, the peptidoglycan of *Mycobacterium abscessus* revealed that 3→3 cross-links were even predominant (from 64 to 74%) depending on the growth phase and bacterial variants (Lavollay *et al.*, 2011). Analysis of the bacterium genome suggested five putative L,D-transpeptidases that await further characterization.

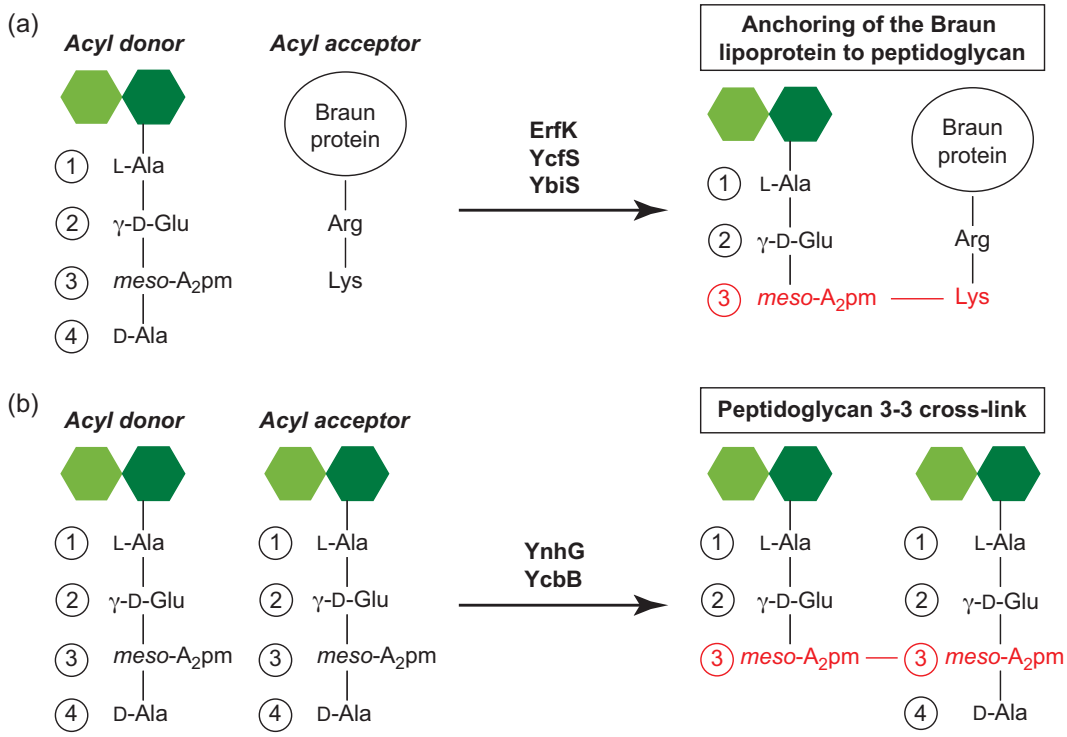


Figure 2.12: Reactions catalyzed by the five L,D-transpeptidases in *E. coli*. **a)** ErfK, YcfS and YbiS enzymes are responsible for the attachment of the Braun lipoprotein to peptidoglycan. **b)** YnhG and YcbB enzymes catalyze the formation of *meso*-A_2pm^3-*meso*-A_2pm^3 peptidoglycan cross-link. The first step that involves the cleavage of the *meso*-A_2pm^3 and D-Ala^4 peptide bond is similar in both reactions, whereas the second step differs by the nature of the acyl acceptor substrate, which is the lysine side-chain of the Braun protein in a), instead of the *meso*-A_2pm^3 side-chain of peptidoglycan acceptor stem in b). Hexagons represent the GlcNAc (light green) and the MurNAc (dark green) molecules.

In *E. coli*, five proteins display sequence similarity with the catalytic domain of LDts from Gram-positive bacteria: ErfK, YcfS, YbiS, YnhG and YbcB. The group of Michel Arthur investigated by multiple chromosomal deletions and genetic complementation the role of the genes encoding for the first four proteins (*erfK, ycfS, ybiS and ynhG*), and showed that three of the putative L,D-transpeptidases (ErfK, YcfS and YbiS) were in fact responsible for the attachment of the Braun lipoprotein to peptidoglycan (Magnet *et al.*, 2007b). This membrane protein is an abundant component of some Gram-negative cell walls, which is covalently bound at its C-terminal end to the peptidoglycan layer and is embedded in the outer membrane by its hydrophobic head, therefore providing structural integrity to the outer membrane. In 2008, the same group demonstrated that the two remaining enzymes, YnhG and YcbB, are in fact the L,D-transpeptidases responsible for the synthesis of the *meso*-A_2pm^3 → *meso*-A_2pm^3 cross-link in the *E. coli* peptidoglycan (Magnet *et al.*, 2008) (figure 2.12).

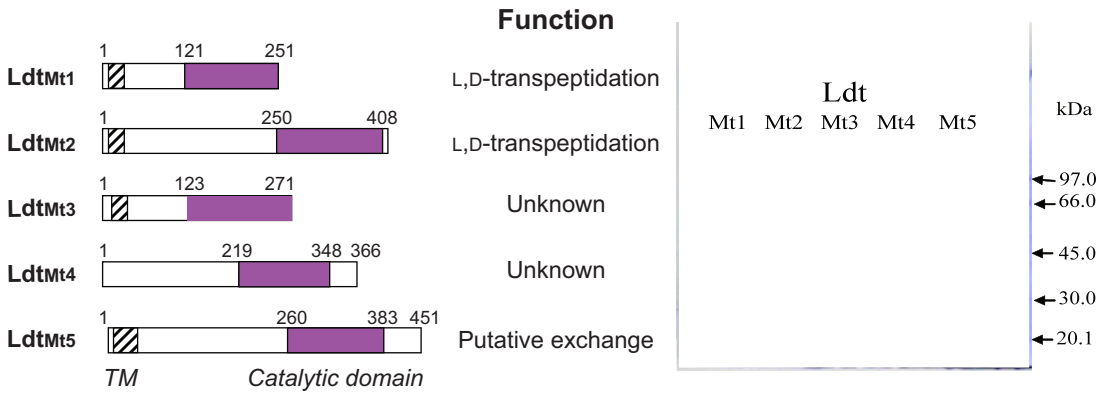
Despite apparent divergence in the previously described enzymatic activities, both reactions involve the same tetrapeptide acyl donor substrate and similar acylenzyme intermediates; only the acyl acceptor

substrates in the second phase of the reaction differ (the side-chain amino group of a C-terminal protein lysine residue in the first place and the side-chain amino group of a peptide stem *meso*-A_2pm in the second case). In both cases nevertheless the term L,D-transpeptidation stands as the first step involves the cleavage of a peptide bond between a L- and a D-amino acid before the formation of a new peptide bond.

The role of L,D-transpeptidation in Mycobacterium tuberculosis

Among all infectious diseases, *M. tuberculosis* is the second cause of mortality, with 8.8 millions new cases each year and 1.1 millions of death in 2010 (WHO, 2011). Furthermore, one third of the worldwide population is infected with *M. tuberculosis* even-though all individuals may not be diagnosed with tuberculosis. The most health-threatening forms of the disease arise when the patient develops a latent infection, in which tuberculosis bacilli survive within the body and avoid the host immune system. Such persistent bacilli can survive several months in the body thanks to the activation of a "dormancy" regulon, even in presence of intensive antibiotic treatment.

In 2002, a gene (Rv0116c) encoding for a member of the active-site cysteine peptidase family was identified in *M. tuberculosis*, that showed a 17-fold upregulation during nutrient starvation, an *in vitro* laboratory model of dormancy (Betts *et al.*, 2002). The structure of *M. tuberculosis* peptidoglycan from a stationary-phase culture in these conditions was thus investigated to evaluate if this gene could spark off the presence of potential 3→3 cross-links. Indeed, the peptidoglycan analysis revealed 80 % of cross-links generated by L,D-transpeptidation (Lavollay *et al.*, 2008). The enzyme responsible for this 3→3 cross-link was subsequently identified as a homologue of Ldt_{fm}, and referred to as Ldt_{Mt1}. In 2010, another LDt was identified in *M. tuberculosis* (Ldt_{Mt2}) and was shown to catalyze 3→3 cross-links so as Ldt_{Mt1} (Gupta *et al.*, 2010). Deletion of ldt_{Mt2} gene led to altered colony morphology, loss of virulence and increased susceptibility to the antibiotic combination of amoxicillin and clavulanate. Recent studies demonstrated that 3→3 cross-links also predominate in *M. tuberculosis* during the exponential growth phase (Kumar *et al.*, 2012).



The group of Michel Arthur identified three additional putative LDts in *M. tuberculosis*, leading to a total number of five LDts in this bacterial strain. A schematic representation of the organization of the different domains (transmembrane and catalytic domains) along the sequence is proposed in figure 2.13, together with their known functional properties.

2.3.2.2 Structure of LDts: the catalytic domain, a shared domain

At the beginning of the present work in october 2009, only two structures of L,D-transpeptidases were known, both determined in the same time frame from X-ray data.

The structure of the *Enterococcus faecium* enzyme (Ldt_{fm}, PDB code: 1ZAT) was solved in 2006 (Biarrotte-Sorin *et al.*, 2006). It concerns in fact a catalytically active fragment of the full-length protein (residues 119 to 466), without the 118 N-terminal residues containing a putative membrane anchor (Mainardi *et al.*, 2005). Residues 119 to 216 were additionally cleaved during the crystallization process and are thus also absent from the crystal structure. The protein crystallizes as one molecule per asymmetric unit and was shown to be monomeric in solution by gel filtration. The final structure of Ldt_{fm} was obtained at a 2.4 Å resolution and displays two distinct domains (figure 2.14a): an elongated N-terminal domain with a new mixed α-β fold (in green), and a ErfK/YbiS/YnhG C-terminal domain (Pfam PF03734, in magenta). The crystallized fragment of the N-terminal domain (residues 217-338) is highly elongated, measuring approximately 65 Å × 20 Å, and consists of nine β-strands and four α-helices. Regarding the structure, it was suggested that, similarly to PBP3 from *S. pneumoniae*, the entire N-terminal domain of Ldt_{fm} (residues 1-338) could act as a pedestal, placing the C-terminal active site in the proximity of nascent peptidoglycan chains. The C-terminal domain (residues 339-466) contains a high content of β-sheets, as well as two conserved residues, a histidine and a cysteine (His^{421} and Cys^{442} in Ldt_{fm}). In the specific case of LDts, this domain is also referred to as the catalytic domain.

The structure of the *Bacillus subtilis* enzyme (Ldt_{Bs}, PDB code: 1Y7M) was solved the same year (Bielnicki *et al.*, 2006). At that time, the protein was referred to as YkuD and its function was unknown, even if it had been previously demonstrated that YkuD transcription was regulated by a factor active during bacterial sporulation (Kodama *et al.*, 2000). It is a year latter that the group of Michel Arthur unambiguously identified *ykud* as the gene encoding the L,D-transpeptidase from *B. subtilis*. The structure of YkuD was solved at a 2.0 Å resolution, with two molecules per asymmetric unit, which form a head-to-tail dimer in the crystal. However, gel filtration indicated that the predominant species was a monomer, suggesting that the dimer is likely to have no functional implication. Furthermore, the dimerization interface involved the residues K117A and G118A that had been mutated for crystallization purposes. The structure revealed two distinct domains (figure 2.14b): an N-terminal LysM domain (in palegreen, residues 1-52) and a C-terminal ErfK/YbiS/YnhG domain (in wheat, residues 53-164). The LysM domain (for Lysin-Motif, Pfam PF01476) has been suggested to be a general peptidoglycan-binding module (Bateman and Bycroft, 2000) as it is found in many of the proteins that are associated to the bacterial cell wall. It is characterized by a βααβ tertiary structure.

The catalytic domains of both Ldt_{fm} and Ldt_{Bs} display large similarities (figure 2.14c and d) as the fold of the two proteins can be superposed (RMSD = 1.21 Å on the 103 Cα backbone atoms in common). They both contain ten β-strands organized in a β-sandwich, with five β-strands on one side, facing the

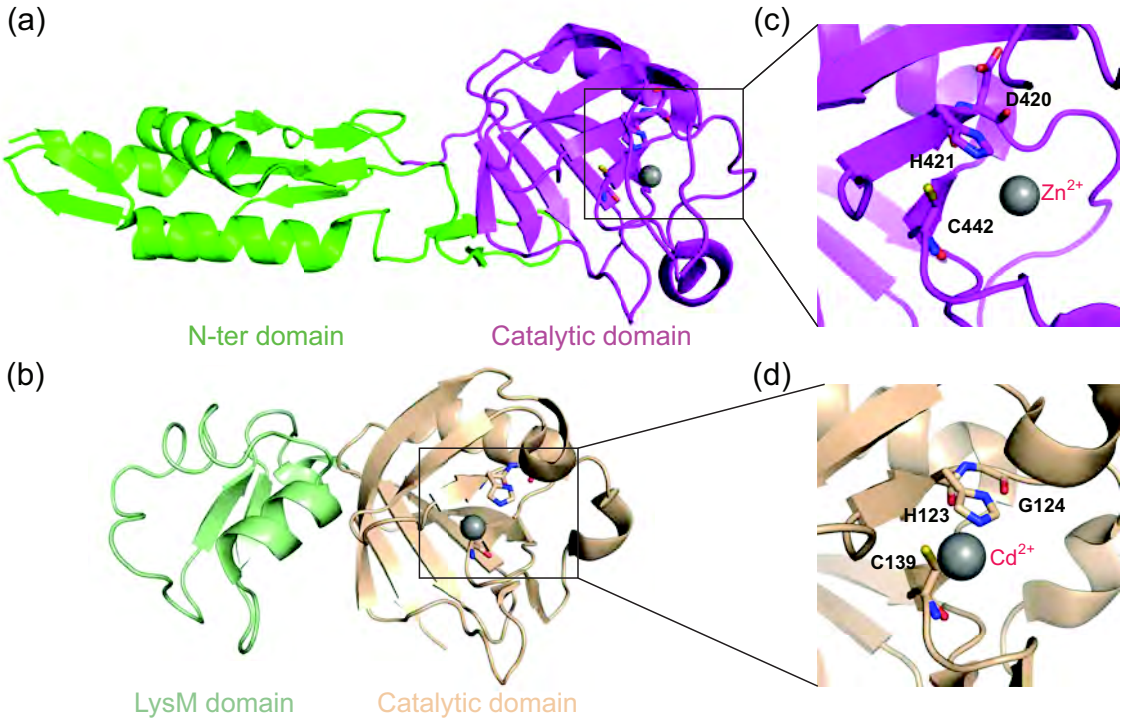


Figure 2.14: X-ray structures of the two L,D-transpeptidases known to date. **(a)** LDt from *E. faecium* (Ldt_{fm}, PDB code: 1ZAT) and **(b)** from *B. subtilis* (Ldt_{Bs}, PDB code: 1Y7M). A closer view into the active site is provided in **(c)** and **(d)** for Ldt_{fm} and Ldt_{Bs}, respectively. The two structures display a similar catalytic domain, containing the active-site cysteine and a conserved histidine residue. In both cases, a divalent metallic cation is present at the active site. Active-site residues including the catalytic cysteine are shown as sticks.

other five in a perpendicular orientation. The main structural difference arises from the long and flexible loop in Ldt_{fm} (residues 389-409), which is replaced by a short and stable loop in Ldt_{Bs} (residues 102-111). This loop might have a functional relevance, as in Ldt_{fm} it is separating the two potential substrate binding paths that lead to the active cysteine (Biarrotte-Sorin *et al.*, 2006). The two conserved residues in the ErfK/YbiS/YnhG domain, His^{421}/Cys^{442} in Ldt_{fm} and His^{123}/Cys^{139} in Ldt_{Bs} are perfectly superposed when the catalytic domains are aligned. This lead the authors to propose functionally relevant catalytic triads to characterize this novel family of cysteine-active enzymes, involving Cys^{442}-His^{421}-Asp^{422} and Cys^{139}-His^{123}-Gly^{124} in Ldt_{fm} and Ldt_{Bs}, respectively. In both cases, the His Nδ1 is suggested to be involved in a hydrogen bond with the oxygen atom from the main chain carbonyl of Asp and Gly in the active enzyme. However, the metallic ions in each structure (Zn^{2+} in Ldt_{fm} and Cd^{2+} in Ldt_{Bs}, respectively) are suggested to disturb the orientation of the active site residues, and specifically the conformation of the cysteine which in Ldt_{Bs} is directly coordinated to the non-physiological cadmium ion (Bielnicki *et al.*, 2006). These structural details point out the need for further structural investigations to extract details on the L,D-transpeptidase active site in the functionally relevant form.

2.3.3 Inhibition of D,D-transpeptidases

2.3.3.1 The β-lactam family

β-lactam antibiotics act specifically on the essential transpeptidation step of peptidoglycan biosynthesis. Today, they are the most wide-spread antibiotics and are used in the treatment of both Gram-positive and Gram-negative bacteria. The most commonly used β-lactam antibiotics include penicillins, cephalosporins, monobactams, carbapenems, and penems. They represent 50% of the marketed drugs of this class (Poole, 2004). β-lactams are characterized by a 4-membered cycle, generally called the β-lactam ring (in red in figure 2.15).

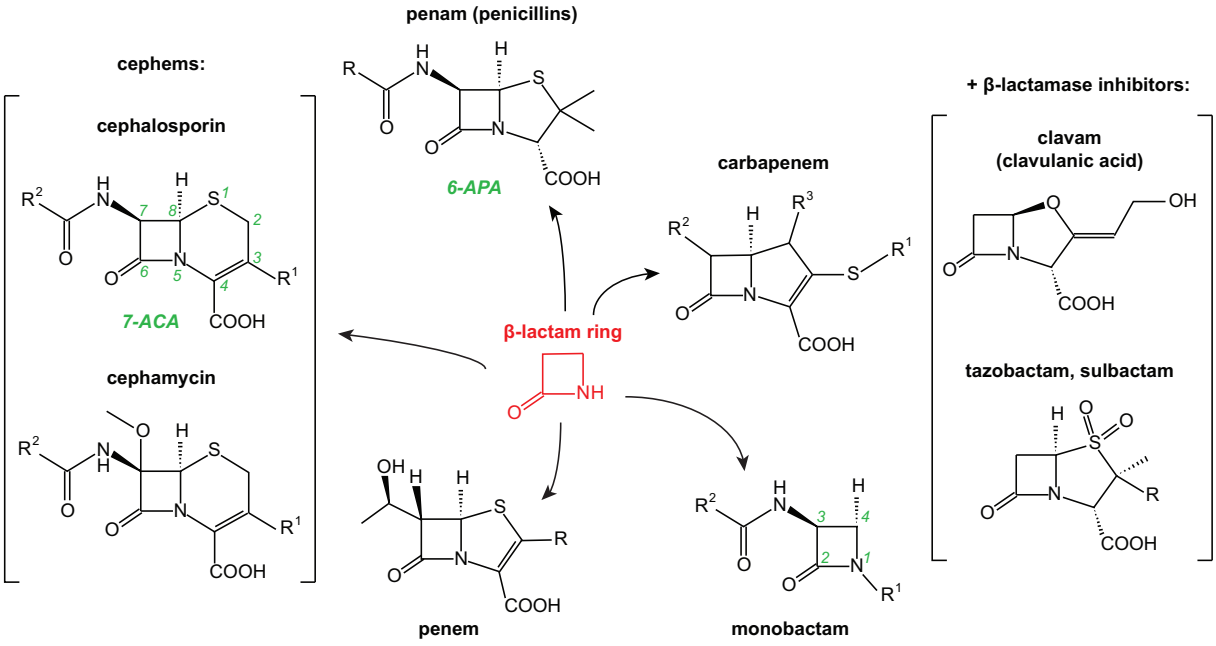


Figure 2.15: Structures of the main families of β-lactam antibiotics. The common motif is the beta-lactam ring represented in the middle. Commercialized beta-lactam antibiotic families include: penam, penem, carbapenem, cephem among which cephalosporin and cephamycin, and monobactam. β-lactamase inhibitors are shown in addition, with the two main representatives: clavam and tazobactam. The common nomenclature of penicillin nucleus and cephalosporin are indicated in green on 6-Aminopenicillanic acid (6-APA) and 7-Aminocephalosporanic acid (7-ACA), respectively. The conventional numbering of bicyclic β-lactams is indicated for cephalosporin, with the position 1 always adjacent to the carbon of β-lactam ring, followed by a clockwise numbering. The numbering of monobactams is different and follows the IUPAC rules, with the nitrogen atom in position 1 and the carbonyl carbon in position 2, as drawn on the monobactam structure.

To provide for the important medical needs in the natural antibiotic **penicillin**, the only β-lactam known during World War II and essential in this context, laboratories in the United States had to develop new methods for mass production. Scientists were able to isolate a novel *Penicillium* strain, *Penicillium chrysogenum*, that produced significantly more penicillin than previously used strains. As the production was spread out to other laboratories, they discovered that different strains, culture conditions, or media could result in the production of different penicillin compounds (Penicillin G, I, F and II) (Kong *et al.*, 2010). With the success of natural β-lactams, scientists were eager to look how organic chemistry could contribute and extend the β-lactams family. Furthermore, with the emergence of resistant strains of *S. aureus* in hospital populations in the early 1950s, it became apparent that new antibiotics were needed. The biosynthesis of penicillin from its precursors, L-cysteine, L-valine and phenylacetic acid opened the

way to produce semi-synthetic penicillin (Sheehan and Henery-Logan, 1959), by modifying the side-chains of the penicillin scaffold also called 6-APA (6-aminopenicillanic acid). In this context methicillin was the first semi-synthetic penicillin introduced for clinical use (Fairbrother and Taylor, 1961). More information on the penicillin family can be found in (Wright, 1999) and (Miller, 2002).

While semi-synthetic compounds were developped, natural sources continued to be explored. **Cephalosporin** was the second natural β-lactam to be discovered after penicillin, isolated from a strain of *Cephalosporium acremomium* (Abraham and Newton, 1961). It yielded a completely new β-lactam family, considering the fact that the 7-aminocephalosporanic acid (7-ACA) scaffold was significantly different from 6-APA (in green in figure 2.15). Using 7-ACA, several generations of cephalosporins were synthesized, some generations displaying broad-spectrum activity. Then came the discovery of **cephamycins**, cephalosporins modified at the C7 position and produced by various species of *Streptomyces* and *Nocardia* (Demain and Elander, 1999). The presence of the methoxy group on the β-lactam ring rendered the molecule more active against anaerobic pathogens and more resistant to Gram-negative β-lactamases, enzymes responsible for the hydrolysis of antibiotics *in vivo*.

In the late 1970s, scientists turned to carbapenems and clavulanic acid, both produced by actinomycetes such as *Streptomyces*. **Clavulanic acids** are part of the β-lactamase inhibitors. The antibiotics of this class exhibit a weak antibiotic activity against a variety of Gram-positive and Gram-negative bacteria; however, they are excellent inhibitors of β-lactamases produced by many bacterial strains resistant to penicillin and cephalosporin. Clavulanic acid is therefore often co-administrated with broad-spectrum penicillins that are susceptible to β-lactamases.

Concerning **carbapenems**, the discovery of thienamycin (Kahan *et al.*, 1979) opened the way to a huge number of new natural carbapenem antibiotics. Thienamycin is still the most potent, larger spectrum and non-toxic natural antibacterial agent ever found. Carbapenems clinically used today are all derived from thienamycin. Their basic five-membered ring differs from the one of penicillin by the presence of a double-bond and the absence of the sulfur atom in position 1. Carbapenems are unique because they are relatively resistant to hydrolysis by most β-lactamases, they can in some cases inhibit these enzymes, and still target PBPs (Papp-Wallace *et al.*, 2011). They also belong to the lone β-lactam class, with penems, that are able to inhibit the L,D-transpeptidases. **Penems** form a recent class of β-lactams, containing only synthetic molecules. They share a similar structure with carbapenems, the main difference being the presence of the sulfur atom in position 1 instead of the carbapenem carbon. Faropenem is the most famous example of this novel class.

Finally, the monocyclic β-lactams, also called **monobactams**, were the last natural β-lactams to be discovered, produced by both actinomycetes and unicellular bacteria including *Pseudomonas* and *Flavobacterium* (Kong *et al.*, 2010). The first member, nocardicin A, was discovered in 1976 and was produced by strains of *Nocardia* and *Streptomyces*. Monobactams are the only β-lactams that do not contain a cycle fused to the β-lactam ring. They differ in the nature of the R^1 and R^2 substituents, as well as in the presence or absence of methoxy group on carbon C3. In many cases, the R^1 substituent consists in SO_3^-, as for example in aztreonam. Most of the natural monobactams have only moderate antibacterial activity, the marketed products are thus all produced by chemical synthesis (Demain and Elander, 1999).

2.3.3.2 PBPs inhibition by β-lactams

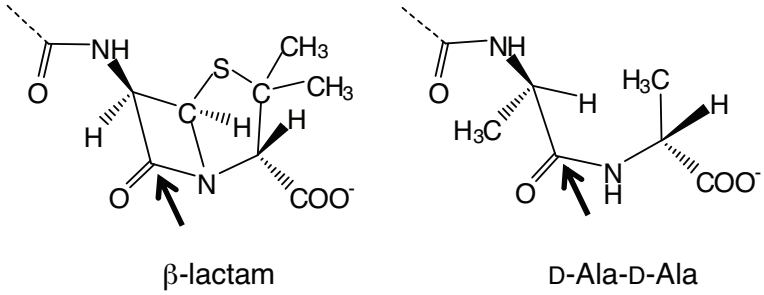


Figure 2.16: Structural similarity between β-lactams and D-Ala-D-Ala from the peptidoglycan pentapeptide.

In 1965, it was suggested that β-lactam antibiotics mimic the conformation of the terminal D-Ala-D-Ala part of the peptidoglycan pentapeptide stem (figure 2.16) (Tipper and Strominger, 1965; Silvaggi *et al.*, 2005). The oxygen of the active serine from the transpeptidase catalytic domain of PBPs attacks the electrophilic carbonyl of the β-lactam ring in a similar fashion as it would attack the carbonyl group of D-Ala^4 of the peptidoglycan pentapeptide stem. In both cases an ester is then formed. In the case of the antibiotics, the formation of the covalent bond with the active serine comes with the opening of the β-lactam ring, which leads to the inactivation of the antibiotics (if it is released). β-lactams are thus known as "suicide substrate". Then, PBPs are not able to perform the transpeptidation anymore because of the steric hindrance due to the presence of the antibiotic in the catalytic site. The performance of β-lactams to inhibit PBP is related to the slow kinetics of the release of the opened beta-lactam ring from the enzyme (Frère *et al.*, 1975, 1976).

2.3.4 Inhibition of L,D-transpeptidases

2.3.4.1 The carbapenem class of β-lactams

Carbapenems and penems are the only β-lactam antibiotics able to inactivate LDts. The carbapenem scaffold consists of a five-membered-cycle attached to the β-lactam ring (figure 2.17). The R_2 substituant is usually a hydroxyethyl (CH_2(OH)-CH_3) extremity, that can harbor different stereochemistry depending on the carbapenem. A methyl is usually found at the R_3 position, except for imipenem and panipenem. Finally, the R_1 side-chain contains the main differences from one carbapenem to another.

Originally developed from thienamycin, a natural product identified in cultures of *Streptomyces cattleya* (Kahan *et al.*, 1979), synthetic production of carbapenems led to over than 80 structures known to date (Demain and Elander, 1999; Papp-Wallace *et al.*, 2011). Imipenem and panipenem belong to the more stable ones as they are less sensitive to natural hydrolysis than the other carbapenems. However, they are more susceptible to deactivation by dehydropeptidases I (DHP-I), an enzyme found in the human kidney. The absence of the methyl group at the R_3 position accounts for this, that can be counteracted by the coadministration with inhibitors of DHP-I, such as cilastatin or betamipron (Norrby *et al.*, 1983). The methyl group on C1 was therefore preferably included in the subsequent synthesized carbapenems. Other characteristics of the carbapenem class are directly related to the antibiotic stability and to the antibacterial potency (Papp-Wallace *et al.*, 2011):

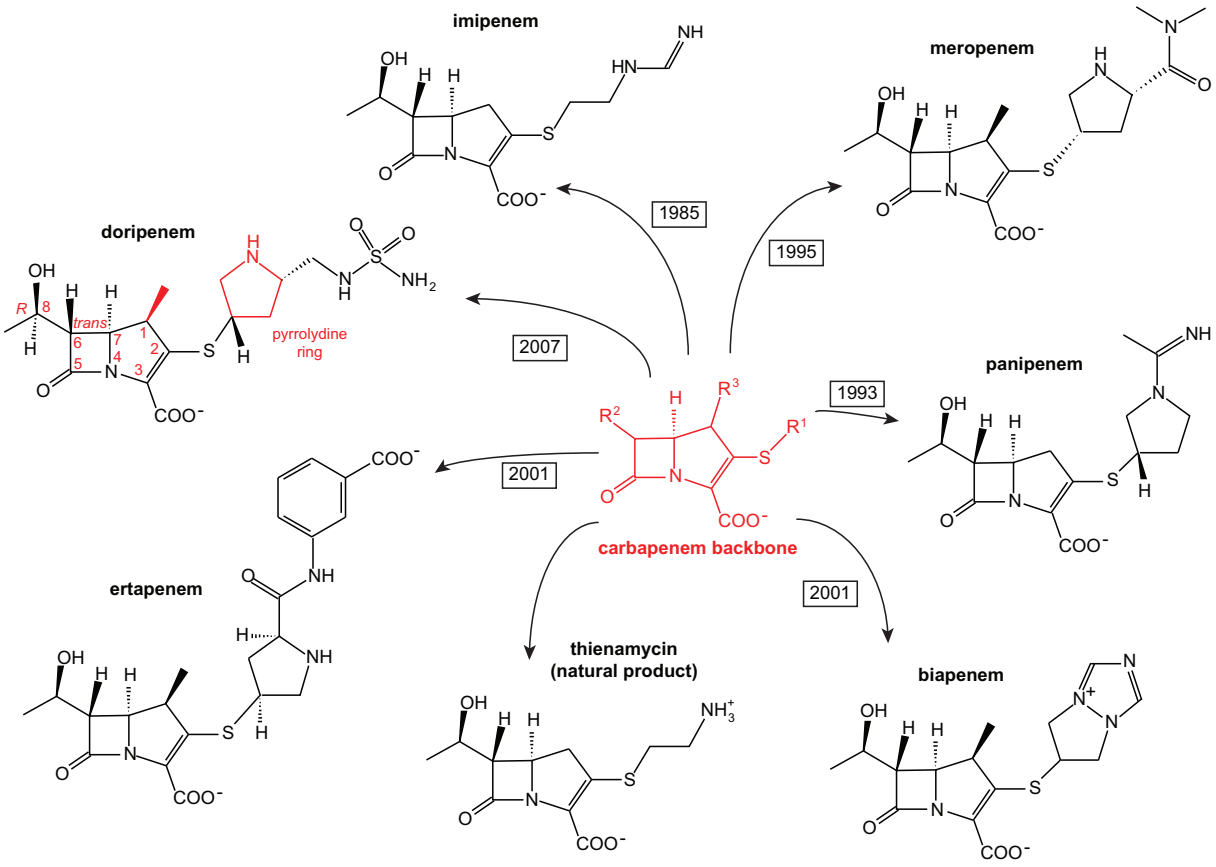


Figure 2.17: Structures of existing carbapenems. The general carbapenem scaffold is shown in the center, surrounded by structures of thienamycin, a natural product identified in 1979 and by the six synthetic most commonly used carbapenems in medecine. Numbers written on the arrows correspond to the year of approval for clinical use (only in Japan in the case of panipenem and biapenem). IUPAC nomenclature is indicated on the doripenem structure, with the characteristics that can impact the stability and potency of the antibiotic.

- The presence of a pyrrolidine ring in R_1 (such as in meropenem, panipenem, doripenem, biapenem and ertapenem) increases stability and broadens the antimicrobial spectrum.
- The presence of a hydroxyethyl group in R_2, that is usually found in the structure of the penem and carbapenem classes, results in stability against β-lactamases.
- This hydroxyethyl can harbor *R* or *S* chirality at the C8 position. The *R* configuration increases the β-lactam's potency. All carbapenems in figure 2.17 harbor this configuration.
- Conversely to penicillin and cephalosporin, the C6-C7 bond of carbapenems is found in a *trans* configuration, which increases the antibiotic's potency and aids in resistance to hydrolysis by β-lactamases.

Among the β-lactams family, carbapenems possess the broadest spectrum of activity and greatest potency against Gram-negative and Gram-positive bacteria. As the other β-lactams, they are not easily diffusible through the bacterial cell wall. In Gram-negative bacteria, they can reach the periplasm thanks to Outer Membrane Protein (OMP). Their mechanism of action is the same as other β-lactams with respect to the acylation of PBPs.

2.3.4.2 LDts inhibition by carbapenems

The inhibition process of LDts by carbapenems involves the formation of a covalent bond between the β-lactam ring carbonyl and the sulfur atom of the active cysteine of the enzyme. Similarly to PBPs with β-lactams, a thioester bond is formed. Yet, in contrast to PBPs, LDts display narrow substrate specificity, as the enzyme inactivation occurs only with β-lactams from the carbapenem class.

The inactivation of LDts by carbapenems was unexpected since transpeptidases with L,D and D,D specificities act on different peptide bonds of the natural substrate, respectively identified as L-Lys^3-D-Ala^4 and D-Ala^4-D-Ala^5 in the donor stem peptide (for the *E. faecium* enzyme), but on the same amide bond of the β-lactam ring. As β-lactams are thought to be structural analogues of the D-Ala^4-D-Ala^5 extremity of the peptidoglycan precursor (Tipper and Strominger, 1965), it is therefore surprising that a molecule such as imipenem could act both as an analogue of the pentapeptide substrate of D,D-transpeptidases and of the tetrapeptide substrate of L,D-transpeptidases. The broadly accepted paradigm that β-lactams inactivate the D,D-transpeptidases because they are structural analogues of a dipeptide comprising two amino acids of the D configuration should therefore be reconsidered, in particular with respect to stereospecificity. Indeed, different β-lactams harbor asymmetric carbons analogous to both D and L amino acid configurations as well as achiral centers and there is no correlation between the stereochemistry of the drugs and their capacity to selectively inactivate transpeptidases of D,D-and L,D-specificities.

2.4 Objectives of the present work

Understanding mechanisms of resistance is critical to accurately detect resistant bacteria in clinical settings, to anticipate and prevent emergence of novel resistance phenotypes, and to design new drugs active against resistant bacteria. The mechanism of inactivation of the classical D,D-transpeptidases by β-lactams and target modification leading to resistance to these drugs has been extensively studied (Sauvage *et al.*, 2008; Zapun *et al.*, 2008). In contrast, little is known about the interaction of carbapenems with L,D-transpeptidases in spite of the high potential of this class of drugs for the treatment of tuberculosis in particular. It is therefore necessary to investigate the mechanism of inactivation of L,D-transpeptidases by this class of β-lactams and to rationalize the specificity for this β-lactam family. Results could also provide insights into the natural transpeptidase or β-lactamase activity of the enzyme at an atomic scale in order to develop new routes for their inactivation.

In this framework, the overall aim of the project in which the present PhD work takes part, is to understand acylation of LDts by β-lactams or by peptidoglycan precursors. The mechanistic study of the transpeptidation reaction with natural substrates is extremely difficult. The enzymatic turnover in the presence of tetrapeptide fragments is too fast to be structurally investigated (Magnet *et al.*, 2007a) and the interaction with product mimetics such as tripeptides is too weak (work of a master student in the lab, O. Buchholz). The acylation by carbapenems was on the contrary considered to be within reach on a biological and technical point of view. Since the reaction of LDts with β-lactams or peptidoglycan precursors is a competitive reaction, the conclusions might be transposable from one substrate to the other.

In order to understand the origin of the specificity of LDts towards carbapenems, the present PhD work chose NMR to explore the structure, thermodynamics and dynamics of different molecules or enzymatic complexes along the acylation reaction pathway, including the apoenzyme, non-covalent complexes with antibiotics and the acylenzyme.

In parallel, the kinetics of the acylation of LDts by different classes of β-lactams was investigated in the group of Michel Arthur, through a combined fluorescence and UV spectroscopy approach (Triboulet *et al.*, 2011). The method exploits the natural fluorescence of tryptophane to study the early stages of the recognition of the antibiotics by LDts (what is believed to correspond to the formation of the Michaelis-Menten complex) and the change in the extinction coefficient upon the β-lactam ring opening and the breakage of the cyclic amide bond. If this approach could not be used on the LDt from *Bacillus subtilis* (Ldt_{Bs}), it offered interesting clues in the case of the enzymes from *Enterococcus faecium* (Ldt_{fm}) (Dubée *et al.*, 2012a) and *Mycobacterium tuberculosis* (Ldt_{Mt1}) (Dubée *et al.*, 2012b). The comparison of the reactivity of different carbapenems with Ldt_{fm} shows that the bulky side-chain of commercial antibiotics have both positive and negative effects in preventing hydrolysis of the acylenzyme and impairing drug binding (Dubée *et al.*, 2012a), while variations in the carbapenem side-chains affected both the binding and acylation steps in Ldt_{Mt1}, ertapenem being the most efficient inactivator. Cephalosporins also formed covalent adducts with Ldt_{Mt1}, although the acylation reaction was 7- to 1,000-fold slower and led to elimination of one of the drug side chains (Dubée *et al.*, 2012b).

The combination of the NMR and kinetics investigations is expected in the long term to provide key atomic-scale information to decipher the inhibition of LDts by β-lactams.

The principles of the NMR methods used to characterize the structure, thermodynamics and dynamics of enzyme-inhibitor complexes, such as the LDt-carbapenem complexes are presented in part II (chapter 3 to 5), together with in-house implementations that were required to solve some of the raised questions.

Three LDts have been studied along the present PhD work, namely the LDt from *B. subtilis* (Ldt_{Bs}), the LDt from *E. faecium* (Ldt_{fm}), and the LDt from *M. tuberculosis* (Ldt_{Mt1}) to a lesser extent, and the results are presented in part III. The complete data interpretation is presented for Ldt_{Bs} in Chapter 6 as a publication, and the on-going work on Ldt_{fm} and Ldt_{Mt1} is detailed in Chapter 7.

Introduction en français

Les bactéries sont des microorganismes procaryotes, de quelques micromètres de longueur. Une de leur caractéristique importante est la paroi cellulaire puisque, étant située à la frontière entre la cellule et son environnement, elle est essentielle pour la survie de la bactérie. Les bactéries peuvent être divisées en deux groupes : Gram-négatif et Gram-positif. Cette séparation est basée sur la différence de composition et de structure de la paroi cellulaire et peut être mise en évidence grâce à la coloration de Gram. Les deux types de bactérie ont une paroi cellulaire composée d'une membrane interne (également appelée membrane cytoplasmique), et d'un nombre variable de couches de peptidoglycane. La paroi des bactéries à Gram-positif (exemple : *Bacillus subtilis*) contient un peptidoglycane très épais et des acides téichoïques, alors que celle des bactéries à Gram-négatif (exemple : *Escherichia coli*) présente un peptidoglycane fin et protégé par une membrane externe. La paroi bactérienne est essentielle pour la bactérie, puisqu'elle lui confère une imperméabilité et une rigidité lui permettant de résister à la pression osmotique cytoplasmique. Le peptidoglycane joue un rôle particulièrement important dans cette résistance, et contribue également à la forme globale de la bactérie. C'est un polymère tridimensionnel composé de chaînes de sucres pontées entre elles par des brins peptidiques. La formation du peptidoglycane a été l'un des sujets d'intérêt majeur de ces dernières décennies, et s'est révélée être d'une grande complexité. Etant une des cibles majeures des antibiotiques, notamment de la plus grande classe connue à ce jour, les β-lactames, le peptidoglycane est demeuré au premier plan de la recherche contre la résistance aux antibiotiques. Même si certaines caractéristiques du peptidoglycane, comme par exemple sa structure tridimensionnelle, restent inconnues, les progrès faits au niveau de la compréhension de sa biosynthèse et de sa composition ont permis quelques avancées pour la recherche de nouveaux antibiotiques.

La biosynthèse du peptidoglycane requiert un certain nombre d'étapes, que l'on peut diviser en trois phases majeures dépendant de leur localisation. Il y a tout d'abord la synthèse du précurseur Uridine Di-Phosphate-*N*-Acetyl-Muramic Acid (UDP-MurNAc) pentapeptide (L-Ala-γ-D-Glu-*meso*-A_2pm-D-Ala-D-Ala), qui se déroule dans l'espace cytoplasmique. Ensuite, ce précurseur est lié au transporteur lipidique, ce qui lui permet d'être accroché à la membrane cytoplasmique, et par un mécanisme encore inconnu à ce jour de passer du cytoplasme au périplasme. Enfin, la polymérisation dans le périplasme des brins peptidiques et glucosidiques permet de former les couches de peptidoglycane grâce à des réactions de transglycosylation et de transpeptidation (voir figures 2.5 et 2.6 pour les différentes étapes de la biosynthèse).

La réaction de transglycosylation permet de polymériser les chaînes de glycane grâce à des glycosyltransférases (GTs). Elle est suivie d'une réaction de transpeptidation réalisée par des D,D-transpeptidases (DDts) qui lient les brins peptidiques sur les chaînes de sucres du peptidoglycane naissant. Ces deux réactions sont catalysées par les Penicillin-Binding Proteins (PBPs). Nous allons nous intéresser plus particulièrement à la réaction de transpeptidation, lors de laquelle les PBPs catalysent la formation d'une liaison peptidique entre la D-Ala en quatrième position du brin donneur du peptidoglycane et l'acide aminé en troisième position du brin accepteur (qui est en général un *meso*-A_2pm pour les bactéries Gram-négatif et une L-Lys pour les bactéries Gram-positif) (figure 2.11) (Goffin and Ghuysen, 1998). Le mécanisme de la réaction implique une attaque nucléophile de la serine du site actif de l'enzyme sur le groupement carbonyle du brin peptidique, formant un premier intermédiaire tétraédrique. L'acylenzyme

est ensuite formée après le départ de l'acide aminé C-terminal D-Ala du substrat. Enfin, l'enzyme est restituée sous sa forme native lorsque le produit N-terminal est relâché.

Pour lutter contre les infections bactériennes, différentes classes d'antibiotiques ont été développées. L'une d'elle, la famille des β-lactames, vise particulièrement les PBPs puisque le noyau β-lactame mime la conformation de la partie terminale du pentapeptide du peptidoglycane D-Ala-D-Ala (Tipper and Strominger, 1965). Les β-lactames forment une liaison ester covalente et irréversible entre l'oxygène de la serine active de la PBP et le groupement carbonyle du noyau β-lactame, entraînant l'ouverture de ce cycle. La présence de l'antibiotique dans le site actif des PBPs ne leur permet donc plus de réaliser la réaction de transpeptidation, ce qui entraîne un affaiblissement voire la mort de la bactérie puisque le peptidoglycane n'est plus formé.

Cependant, les bactéries ont développé des mécanismes de résistance aux antibiotiques. Dans les bactéries à Gram-négatif, les pompes à efflux, qui sont des protéines de transport, permettent d'exporter les composés chimiques toxiques pour la bactérie, tels les antibiotiques, hors de celle-ci. Un autre mécanisme de résistance retrouvé dans tous les types de bactérie, met en oeuvre l'activation des β-lactamases. Ces enzymes provoquent l'hydrolyse d'un grand nombre d'antibiotiques à spectre étendu, libérant ainsi la voie aux PBPs et leur permettant de poursuivre leur rôle dans la biosynthèse du peptidoglycane. C'est d'ailleurs la raison pour laquelle les β-lactames (notamment les pénicillines) sont généralement co-administrés avec de l'acide clavulanique, inhibiteur irréversible de la β-lactamase. Cependant, l'adaptation de cette enzyme, par les nombreuses mutations apparues ces dernières années dans son génome, est un facteur inquiétant qui doit être pris en compte pour la conception des nouveaux antibiotiques. Un autre mécanisme de résistance se trouve dans l'expression de PBPs de faible affinité, qui parviennent à maintenir la réaction de transpeptidation malgré la présence des β-lactames. Enfin, l'activation d'un mécanisme de transpeptidation différent de celui des PBPs, et effectué par une nouvelle classe d'enzyme, les L,D-transpeptidases (LDts), a été identifié dans plusieurs souches bactériennes. Par exemple, il a été montré que 100% des muropeptides de la souche mutante de laboratoire M512 de *E. faecium* hautement résistante à l'ampicilline (l'un des β-lactames le plus utilisé) sont pontés par des LDts. De plus, ce nouveau mécanisme représente un intérêt particulier dans la recherche contre la tuberculose, puisqu'il a été identifié dans plusieurs souches non réplicatives de *Mycobacterium tuberculosis* et serait impliqué dans 80% des pontages.

Au lieu du pontage 4→3 classique effectué par les PBPs entre la D-Ala^4 du brin donneur et le troisième résidu du brin accepteur du peptidoglycane, les LDts forment un pontage 3→3 entre les résidus en troisième position des brins donneur et accepteur. Ce mécanisme est possible suite à l'action des carboxypeptidases, qui clivent la D-Ala^5 du pentapeptide, générant un donneur tetrapeptide, substrat naturel des LDts (voir figure 2.11). Ainsi, même lorsque les PBPs sont inhibées par les β-lactames, les LDts peuvent contourner le problème en formant un peptidoglycane de composition légèrement différente. De la même manière que les PBPs utilisent leur serine active pour aller attaquer le groupement carbonyle du brin peptidique, chez les LDts c'est une cystéine qui joue ce rôle. Le mécanisme est très similaire, bien que les deux types d'enzymes différent assez largement en terme de séquence et de structure.

Les LDts sont résistantes à la plupart des β-lactames, excepté à la famille des carbapénèmes. Bien que le mécanisme exact de l'inhibition des LDts par les carbapénèmes soit inexpliqué, il a été démontré qu'une liaison covalente se forme entre leur noyau β-lactame et le soufre de la cystéine de l'enzyme. Au début de mes travaux de thèse, seules deux structures de LDts étaient connues, contenant toutes deux

des ions métalliques au sein de leur site actif. Ces ions sont des artefacts dus à la cristallisation des enzymes, cependant ils entraînent une modification de l'orientation des résidus catalytiques, y compris de la cystéine active.

L'objectif de ce travail est de comprendre le mécanisme d'inhibition des LDts par les β-lactames, et d'étudier cette réaction dans différentes souches bactériennes afin de comprendre la spécificité des différents β-lactames, et, de manière plus générale, d'apporter des éléments pour la conception de nouveaux antibiotiques efficaces contre les LDts. Pour cela, des études structurales et dynamiques ont été faites par RMN sur les LDts de *Bacillus subtilis* (Ldt_{Bs}) et *Enterococcus faecium* (Ldt_{fm}), dans leur forme libre et leur forme acylée par des carbapénèmes. Le complexe non-covalent, intermédiaire réactionnel, a également été étudié, afin de déterminer si celui-ci tient un rôle dans la reconnaissance de l'antibiotique et la formation de la liaison covalente. Les méthodes RMN utilisées pour les différentes études de cette thèse seront développées dans la partie 2, avec notamment la détermination des états de protonation et des pKas des résidus du site actif des deux enzymes, ainsi que la mesure de constantes d'affinité et la caractérisation de la dynamique. Enfin, la troisième et dernière partie sera l'occasion d'apporter une étude comparative sur les mécanismes d'inhibition (ou certains de leurs éléments) par les β-lactames de différentes souches bactériennes étudiées.

Part II

Implementation of NMR tools for the study of non-covalent and covalent protein complexes: structural, thermodynamical, and dynamics aspects

Introduction to second part

In living systems, protein-ligand interactions are involved in a wide range of biological processes, including drug binding, antigen recognition, and enzyme-inhibitors interactions (Craik and Wilce, 1997). The ligand can be an organic molecule, a peptide, an oligonucleotide, or a carbohydrate. The recognition of a unique binding interface by the ligand is usually the key to functional efficiency of the enzyme. The identification of dedicated interaction networks may lead to design ligands with enhanced binding properties. In the field of new antibiotics discovery, a search for drugs with a higher specific recognition, a larger potency towards the target, or more resistance to hydrolysis is usually critical. Nevertheless rational design can be hampered by additional phenomena that complicate the simple lock-and-key recognition based model.

It is thus important to investigate the interaction between a biological system and a ligand by a variety of biophysical methods. Before to describe the NMR performances with that respect, two widely used techniques are briefly introduced: Isothemal Titration Calorimetry (ITC) and fluorescence spectroscopy.

Isothermal Calorimetry (ITC) is a physico-chemical technique that can provide the complete thermodynamic profile (stoichiometry, affinity constant, enthalpy and entropy) of a ligand-receptor interaction in a single experiment (Jelesarov and Bosshard, 1999). Mainly used for protein-ligand interactions, ITC is based on the titration of two binding partners at constant temperature, through a series of injections of the titrant (usually the ligand) onto the titrate (usually the macromolecule). The heat released or absorbed in the sample cell containing the titrant is measured with respect to a reference buffer cell (Duff *et al.*, 2011). Curve fitting of experimental data combined with the known quantities of the introduced binding partners gives access to the thermodynamical parameters. Advantages of ITC lie in its ability to detect multiple binding events with different characteristics and to be sensitive to a wide range of binding affinity (K_D in the nanomolar to micromolar range). Some of the limits of the technique stands in the consuming experiment-time (typically one day for a complete titration, to repeat at least three times for a rigorous analysis), and the temperature required for the study (higher than room temperature).

Fluorescence spectroscopy has also been extensively used to characterize ligand binding to proteins. Advantages of fluorescence-based methods stand in their high sensitivity to local environmental changes and in the two-at-a-time qualitative and quantitative description of the interaction (Bhattacharyya *et al.*, 2010). Fluorescence methods can provide information on binding mechanism, the number of binding sites, and thermodynamic as well as kinetics parameters. One of the limitations concerns the requirement for a fluorophore. This fluorophore can be either an intrinsic probe (naturally-fluorescent biomolecules, such as Tyr and Trp amino acids, or small molecule cofactors such as flavines), or an extrinsic probe, which is covalently or non-covalently bound to the studied biomolecule. However, in the case of proteins containing several tryptophans, it is particularly difficult to specifically assign the fluorescence change. This methods allows to investigate complexes with dissociations constants in the pM to microM range. In addition, due to the time-scale of optical transitions, fluorescence is particularly adapted to study complexation kinetics, as in the case of the interaction of β-lactam antibiotics with $\mathrm{Ldt_{fm}}$ (Triboulet *et al.*, 2011).

ITC, fluorescence, and the other techniques used to study protein-ligand interactions are nevertheless

not able to localize the ligand position, and can only provide global information on the interaction. In contrast, the power of NMR resides in the characterization at an atomic level of the interaction, enabling the localization of the binding site(s). NMR is sensitive to a wide range of affinities, and is especially used in the detection of low-affinity complexes (micromolar to millimolar range) (Carlomagno, 2005), although the other methods are often limited towards this side of the affinity scale.

The NMR methods used to determine thermodynamic parameters of the interaction of LDts with β-lactam antibiotics are detailed in Chapter 4.

Then, NMR is the only method with X-ray crystallography that enables to solve the three-dimensional structure of biomolecules, providing a visual framework for understanding their mechanism of action. More recently, the technique of electron microscopy was used for the determination of protein structures that could reach the atomic resolution (Chiu, 1993; Stowell *et al.*, 1998); however it is more often used as a combined technique with X-ray crystallography and is useful for very large assemblies. The principle of X-ray crystallography is based on electron-diffraction and requires the obtention of a protein crystal (Wagner *et al.*, 1992). Advantage of this technique resides in the access to large proteins, one of the best example being the structure determination of the ribosome 80S subunit at 4.15 Å resolution (Ben-Shem *et al.*, 2010), while NMR is limited to 40-50 kDa proteins. Furthermore, density maps obtained after a Fourier transform of the diffraction pattern directly give the accurate atomic positions while the NMR structure determination protocol is mainly based on indirect information. However, the main difficulty encountered with X-ray crystallography is the obtention of the protein crystal, a particularly difficult task for membrane proteins or highly soluble proteins. In addition, the requirement for adjuvant to induce the crystallization process may lead to artifacts, as in the case of Ldt_{Bs} and Ldt_{fm} crystal structures where a divalent ion disturbs the active site. NMR is therefore particularly interesting regarding this point of view, as the protein is usually found in its native conformation. With regard to enzyme/inhibitor complexes, conditions of the study are a key factor, as interactions might be really sensitive to environment.

The NMR methods used to determine the structures of LDt apoenzymes and of carbapenems-acylated LDts are detailed in Chapter 3.

Finally, as structure only provides a static point of view, the enzyme/inhibitor interaction should also be investigated on a dynamics point of view. Yet, only a few techniques are available for this purpose, among which neutron scattering (Gabel *et al.*, 2002), molecular dynamics and NMR. Neutron scattering only provides global dynamics, and molecular dynamics is rather time-consuming and usually only provides an in-vacuum description, which is problematic for solvent exposed catalytic sites as in the case of LDts. NMR is therefore a tool of choice by providing a per-residue description of the dynamics parameters on a large time-scale.

The NMR methods used to determine the dynamics parameters of LDts are detailed in Chapter 5.

The principle of NMR spectroscopy consists in the property of many elements to have a nuclear magnetic moment. A nucleus is observable by NMR only if it presents a non-null spin I. If the nuclear spin is superior to $\frac{1}{2}$ (quadrupolar nucleus), relaxation properties of the nucleus are affected and the lines become intrinsically broader, hampering detection. Consequently, the most interesting nuclei to study by NMR are those with spins = $\frac{1}{2}$, which concerns stable isotopes such as ^{1}H, ^{13}C, and ^{15}N, which are

particularly important in biological macromolecules (Table 2.3).

Nucleus	Spin I	Natural abondance (%)	Gyromagnetic ratio γ ($\times$ 10^7 rad.(T.s)$^{-1}$)
^{1}H	1/2	99.98	26.752
^{2}H	1	0.02	4.107
^{12}C	0	98.89	
^{13}C	1/2	1.11	6.728
^{14}N	1	99.64	19.34
^{15}N	1/2	0.36	-2.712
^{16}O	0	99.96	
^{17}O	5/2	0.04	-3.628
^{19}F	1/2	100.00	25.181
^{23}Na	3/2	100.00	7.080
^{31}P	1/2	100.00	10.841
^{33}S	3/2	0.76	5.934

Table 2.3: Nuclear properties of different isotopes for nuclei present within biological macromolecules. Isotopes with spin = 0 are not observable by NMR. This explains why ^{13}C and ^{15}N labeling are required to study macromolecules by NMR.

The NMR signal frequency is extremely sensitive to changes in the nucleus environment. Internuclear scalar and dipolar interactions are also affected leading to modifications in the signal shape and intensity. The resulting NMR spectra therefore provide rewarding information on the environment of each residue as well as insights into the interaction between a target macromolecule and a ligand (Maurer, 2005).

The versatility of the NMR method can thus be used to determine the structure of the enzymes whether in their free state or in complex with a ligand, to characterize the thermodynamics relevant to functional aspects (K_D and Ka determination) and to extract individual or collective motion contributions. This is the reason why NMR was extensively used for the study of LDts.

Chapter 3

Structural approach by NMR

Since the first three-dimensional structure was solved by NMR in Wüthrich group (Williamson *et al.*, 1985), great progresses have been made, that have significantly improved the performance of NMR in the structure determination of larger macromolecular assemblies in reduced time frames. First, technological progresses consisted in increasing NMR resolution and sensitivity, the weaknesses of the technique, through the development of high-resolution spectrometers at higher fields (until 1 Gigahertz to date) and of cryogenic probes. To these advances may be added the emergence of uniform (^{15}N, ^{13}C, ^{2}H) and more recently specific labeling methods (Tugarinov *et al.*, 2006; Gans *et al.*, 2010; Ayala *et al.*, 2012). Finally, the spectroscopic advances, including the development of 3D and 4D NMR sequences (Cavanagh *et al.*, 1995; Sattler *et al.*, 1999; Kanelis *et al.*, 2001), the setting up of the Transverse Relaxation Optimized SpectroscopY (TROSY) experiment (Pervushin *et al.*, 1997; Zhu *et al.*, 2004), and the use of Residual Dipolar Couplings (RDCs) (Tolman *et al.*, 1995; Tjandra and Bax, 1997) significantly improved the resolution of the determined structures and extended the size limit of the macromolecules studied by NMR. To date, the largest single chain protein for which structure has been solved by solution NMR is a 723 residues enzyme from *E. coli*, malate synthase G (MSG) (Tugarinov and Kay, 2003; Grishaev *et al.*, 2008). Using these latest methodological developments, partial structural and dynamics information could be obtained on even larger supramolecular machineries such as a 900 kDa Gro chaperonin (Fiaux *et al.*, 2002) or a 850 kDa proteasome complex (Religa *et al.*, 2011).

The present chapter focuses on the NMR approaches and methods used for the structure determination of LDts, including backbone and side-chain assignment, extraction of structural restraints and semi-automated structure calculation. Additional strategies specific to the study of non-covalent or covalent complexes with β-lactam inhibitors were implemented : assignment of small unlabeled molecules and extraction of intermolecular restraints through chemical shift or NOE information. The latter strategies allowed in particular the structural characterization presented in part III for the LDt acylenzymes after reaction with carbapenems, imipenem for the *B. subtilis* LDt and ertapenem for the *E. faecium* LDt, respectively.

3.1 Protein NMR structure determination

3.1.1 Strategies for protein assignment

The first step towards any structure determination by Nuclear Magnetic Resonance is the so-called "assignment" step. It consists in the identification of each NMR signal to its corresponding nucleus in the protein. This step is determinant as from the exactitude and completeness of the protein assignment will depend the quality of the structure and the swiftness to which it will be solved. The first NMR structure, obtained in 1985 on a 53-residues unlabeled protein was done only using the protons chemical shifts, lone NMR-active nuclei at natural abundance before the advent of isotopic labeling techniques, and with

the sequential assignment method based on the Nuclear Overhauser Effect (NOE) (Williamson *et al.*, 1985). However, for proteins larger than 10 kDa, the use of heteronuclear NMR experiments is required (Breukels *et al.*, 2011). With molecular weights lower than 30 kDa, the assignment of Ldt_{Bs} and Ldt_{fm} did not require deuteration. An efficient protocol adapted to ^{13}C,^{15}N-uniformly labeled protein samples was thus adopted.

3.1.1.1 Sequential backbone assignment

The interpretation and assignment of NMR spectra generally starts with the backbone assignment, including N, H^N, C^α, C' and C^β atoms. With the exception of the proline residue and the side-chain of Asn and Gln, each amino-acid contains a unique amide NH group. The spectral signature for this group is nevertheless not unique for a given amino acid type and position in the primary sequence. When this signature is completed by the signature of backbone carbon, amino acid type can be in some cases deduced, but it remains insufficient for specific assignment. The latter strategy thus requires triple resonance experiments that will connect nuclei from a given residue to its sequential neighbors. These experiments exploit scalar J-couplings between bonded nuclei (see figure 3.1 for J couplings values) to transfer magnetization from one nucleus to another. The time-scale required to transfer magnetization through this interaction is typically in the order of $1/J$ and must be faster than the T_2 relaxation time constant of the starting nucleus. The advantage of heteronuclear over homonuclear experiments clearly appears when heteronuclear J values presented in figure 3.1 are compared to the homonuclear $^3J_{H-H}$ coupling constant (< 15 Hz). The most commonly used triple resonance NMR experiments for protein backbone assignment are summarized in Table 3.1.

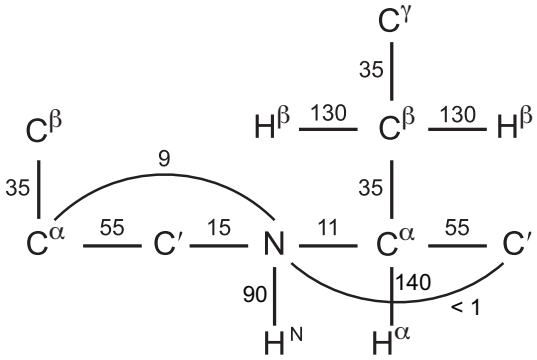


Figure 3.1: Typical values of ^1J and ^2J scalar coupling constants for a ^{15}N-^{13}C-labeled protein. Values are given in Hertz. Adapted from (Cavanagh *et al.*, 1995).

The acronym of each experiment gives precision on the nuclei for which resonance frequency is collected (in plain letter) and on the nuclei that are passively involved in the coherence transfer (within parentheses, or written in lowercase). The latter are thus absent from the spectrum. As an example, the HNCA experiment correlates $^1H^N$ and ^{15}N resonances to the $^{13}C\alpha$ of the same residue *i* and to the $^{13}C\alpha$ of the preceding residue (*i*-1), whereas the HN(CO)CA experiment correlates $^1H^N$ and ^{15}N resonances only to the $^{13}C\alpha$ of the (*i*-1) residue via the non-edited carbonyl carbon. HNCA and HN(CO)CA are therefore complementary and can be interpreted in pairs. Similar conclusions apply to the HNCO / HN(CA)CO pair (which correlates ^1H_N, ^{15}N and carbonyl carbon resonances) and the HNCACB / HN(CO)CACB pair (which correlates ^1H_N, ^{15}N and Cα and Cβ aliphatic carbon resonances) (Table 3.1). The common ^1H_N and ^{15}N dimensions then allow to connect the different pairs of 3D experiments together.

Experiment	Nuclei observed	Relative S/N (in %)
HNCO	H_i, N_i, C'_{i-1}	100
HN(CA)CO	H_i, N_i, C'_i, C'_{i-1}	i: 13 and i-1: 4
HN(CO)CA	H_i, N_i, C^{α}_{i-1}	71
HNCA	H_i, N_i, C^{α}_i, C^{α}_{i-1}	i: 50 and i-1: 15
HN(CO)CACB	H_i, N_i, C^{α}_{i-1}, C^{β}_{i-1}	α: 13 and β: 9
HNCACB	H_i, N_i, C^{α}_i, C^{β}_i, C^{α}_{i-1}, C^{β}_{i-1}	α_i: 4 and β_i: 1.7
		α_{i-1}: 1.3 and β_{i-1}: 0.5

Table 3.1: Summary of the most used triple resonance NMR experiments in protein backbone assignment, often used in pairs. The Signal/Noise (S/N) ratio is relative to the HNCO experiment, estimated for a 20 kDa protein. HN(CO)CACB offers the same correlations than the CBCA(CO)NH, so as HNCACB and CBCANH. Table adapted from (Sattler *et al.*, 1999).

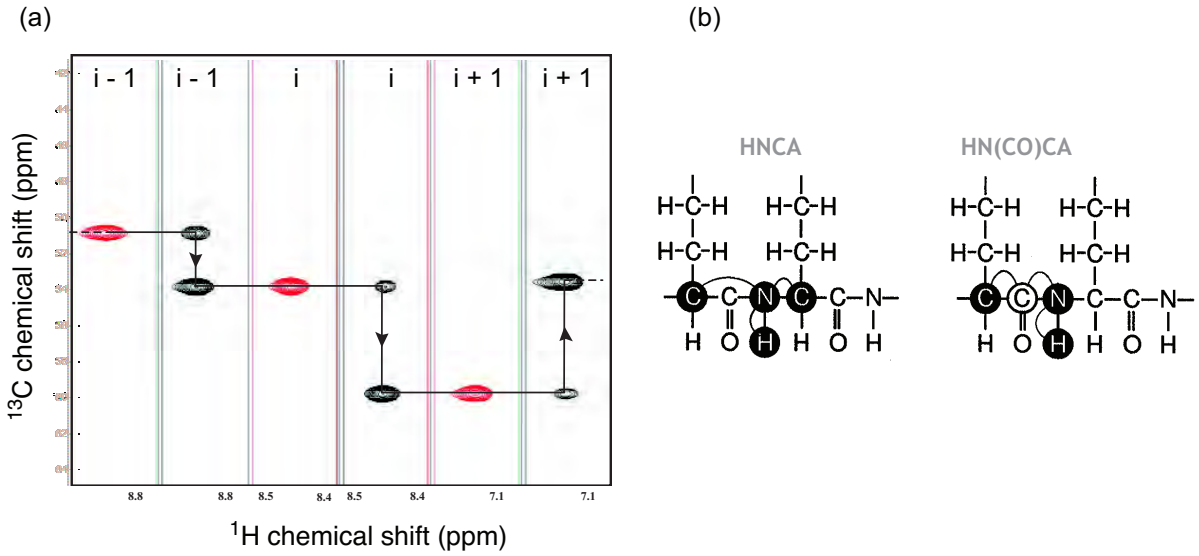


Figure 3.2: Pair of 3D triple resonance experiments used for the assignment of protein backbone resonances. **a)** Example of strips extracted from ^{15}N-planes of HNCA (in black) and HN(CO)CA (in red) spectra. Arrows indicate the assignment path from *i-1* to *i+1* residue. **b)** Nuclei involved in coherence transfer in HN(CO)CA and HNCA experiments. The resonance frequency of nuclei with filled circles is edited on the contrary to nuclei with open circles..

^{1}H-^{13}C strips are extracted from each of the full 3D spectra at a given ^{15}N-frequency and are aligned to allow a sequential walk as demonstrated in figure 3.2. Specific assignments are then achieved through the identification of characteristic and easily recognizable chemical shift patterns of some residues, such as the C^{α} of glycines, or C^{β} of threonines, serines and alanines, with respective average values of 45.36 ppm, 69.70 ppm, 63.79 ppm, and 19.00 ppm according to the Biological Magnetic Resonance Data Bank (BMRB)). It was shown that the combined C^{α} / C^{β} chemical shifts of three or four adjacent residues are sufficient for mapping this fragment on the protein primary sequence (Grzesiek and Bax, 1993).

For small ($<$ 15 kDa) folded proteins with nicely dispersed signals in the ^{1}H-^{15}N-Heteronuclear Single-Quantum Coherence (HSQC), the lone HNCACB experiment can be sufficient to do the sequential assignment. The intra- and inter- signals can be differentiated on the basis of the intensity, even if one should be careful as peak intensity also depends on residue-specific parameters, essentially relaxation. For larger proteins or α-helicoidal proteins (for which the signals are very similar in term of ^{1}H and ^{15}N chemical shifts), the combined use of several pairs of triple resonance experiments presented on Table 3.1 is often required in order to limit errors and to get rid of ambiguities. In the case of the 20 kDa Ldt_{Bs} a combination of HNCACB/HN(CO)CACB and HNCO/HN(CA)CO offered essentially complete assignments.

In the case of proteins with molecular weight lower than 20 kDa like the catalytic domain of Ldt_{fm}, backbone assignments can also be facilitated thanks to a HADamard-encoded AMino-ACid-type- editing (HADAMAC) experiment (Lescop *et al.*, 2008), that merges the 20 possible residues found in a protein into seven groups, using a single fast 2D experiment (figure 3.3). Each ^{1}H-^{15}N correlation peak is sorted according to the amino-acid group type of the preceding residue (see also Feuerstein *et al.* (2012) for the intra-residue version). The sorting of the 20 natural amino acids is based on the Cα or Cβ frequencies of certain amino acids and the spin-coupling topologies, *i. e.* the number of attached protons to each of the Cα or Cβ carbons, and the number and type (aliphatic, aromatic, or carbonyl) of carbons at the γ position. It also takes advantage of the Hadamard approach that utilizes different combinations of band-selective frequency to excite nuclei of interest (Kupce and Freeman, 2003). This method thus allows an easy and fast identification of the residue-type (intra or preceding residue) for each ^{1}H-^{15}N correlation peak, that is particularly helpful in crowded regions of the spectrum containing accidental spectral overlap. When coupled to a HNCACB experiment, the HADAMAC experiment affords efficient assignment protocol and allows a comfortable reduction in spectrometer and data-processing times.

Backbone assignment offers information on the protein secondary structure elements in proteins through chemical shifts. The "Chemical shift index" method is used for determining the identity, extent, and location of secondary structure elements in proteins based on the chemical shift difference between the observed chemical shift and the random coil value. The consensus sign of this difference for all backbone nuclei of a same residue allows to establish its occurrence in and α-helix or a β-strand. (Wishart and Sykes, 1994).

If this analysis offers the secondary structure elements, their relative position remains unsolved and the fold of the protein can not be established at this stage. Distances and/or angles between these elements have to be added. Nevertheless the main contacts occur through the side-chain protons, thus requiring their assignment.

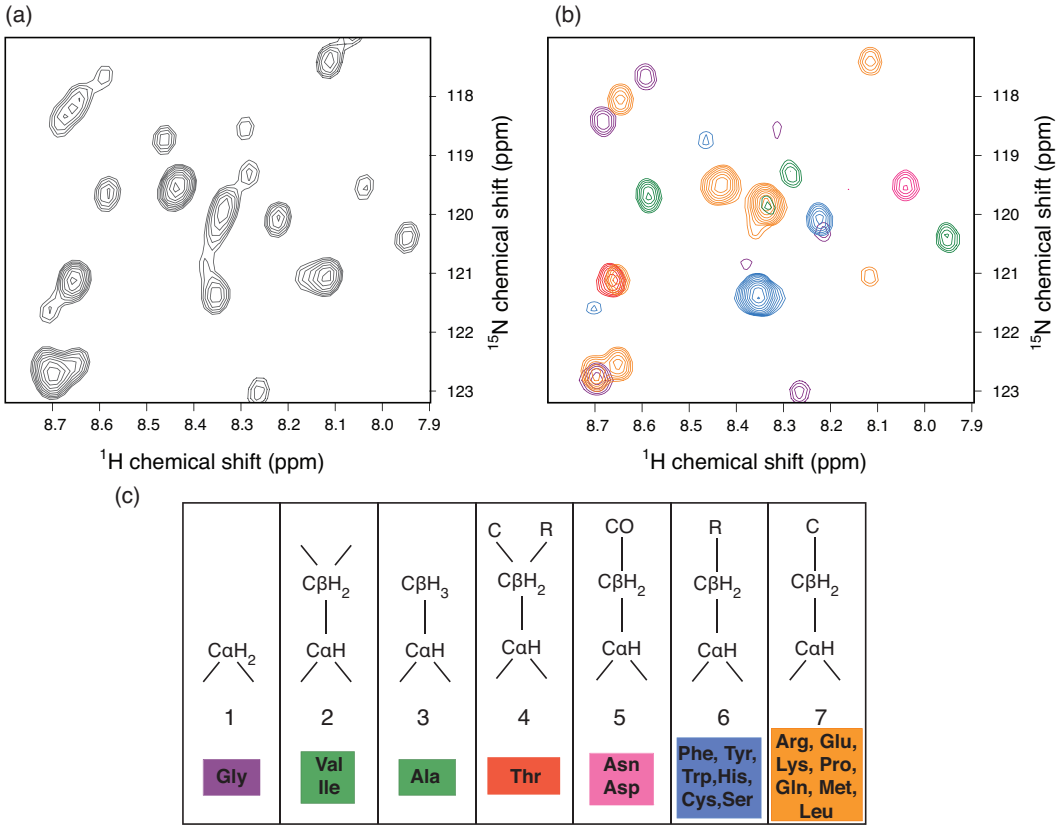


Figure 3.3: Improvement of NMR assignment fastness with hadamard spectroscopy. a) Standard ^{1}H-^{15}N correlation spectrum and **b)** ^{1}H-^{15}N-HSQC correlation spectrum using hadamard technique. Each color correspond to one of the seven amino acid groups for the preceding residue, detailed in panel **c)**. Both spectra were recorded on the same sample of Ldt$_{fm}$ at pH 6.4, 1 mM and 25°C.

3.1.1.2 Side-chain assignment

Despite a large chemical shift dispersion, especially in the carbon dimension (figure 3.4), ^{1}H-^{13}C correlation 2D spectra show major overlaps and are inefficient to provide unambiguous side-chain assignments.

The 3D experiments used to assign aliphatic protons and carbons along the amino acid side-chains were firstly based on the coherence transfer through homonuclear $^{2}J_{H-H}$ and $^{3}J_{H-H}$ scalar couplings, as in the ^{13}C-edited 3D TOCSY (for TOtal Correlation SpectroscopY)-HSQC. However, these couplings are small (4-10 Hz), and lead to significant signal loss during the coherence transfer when the molecular weight of the protein increases. The following approaches then utilized the well-resolved one-bond $^{1}J_{C-C}$ (30-60 Hz) scalar couplings to transfer coherence.

Experiments were thus designed in the literature that included a ^{13}C-^{13}C TOCSY evolution period to allow the coherence transfer along the aliphatic ($^{1}J_{C-C}$ = 35 Hz) or aromatic ($^{1}J_{C-C}$ = 60 Hz) C-C side-chains (Bax *et al.*, 1990). In these experiments, the extent of transfer from one end to the other of the side-chain is dictated by the length of the isotropic mixing and the T_2 (transverse) relaxation time constants of individual ^{13}C nuclei. Long mixing times favor long-range transfers, while large protein molecular weights and associated fast relaxation processes hamper them. In the case of Ldt$_{Bs}$ (20 kDa) and Ldt$_{fm}$ (15 kDa) a mixing time of 140 ms was selected for aliphatic side-chains.

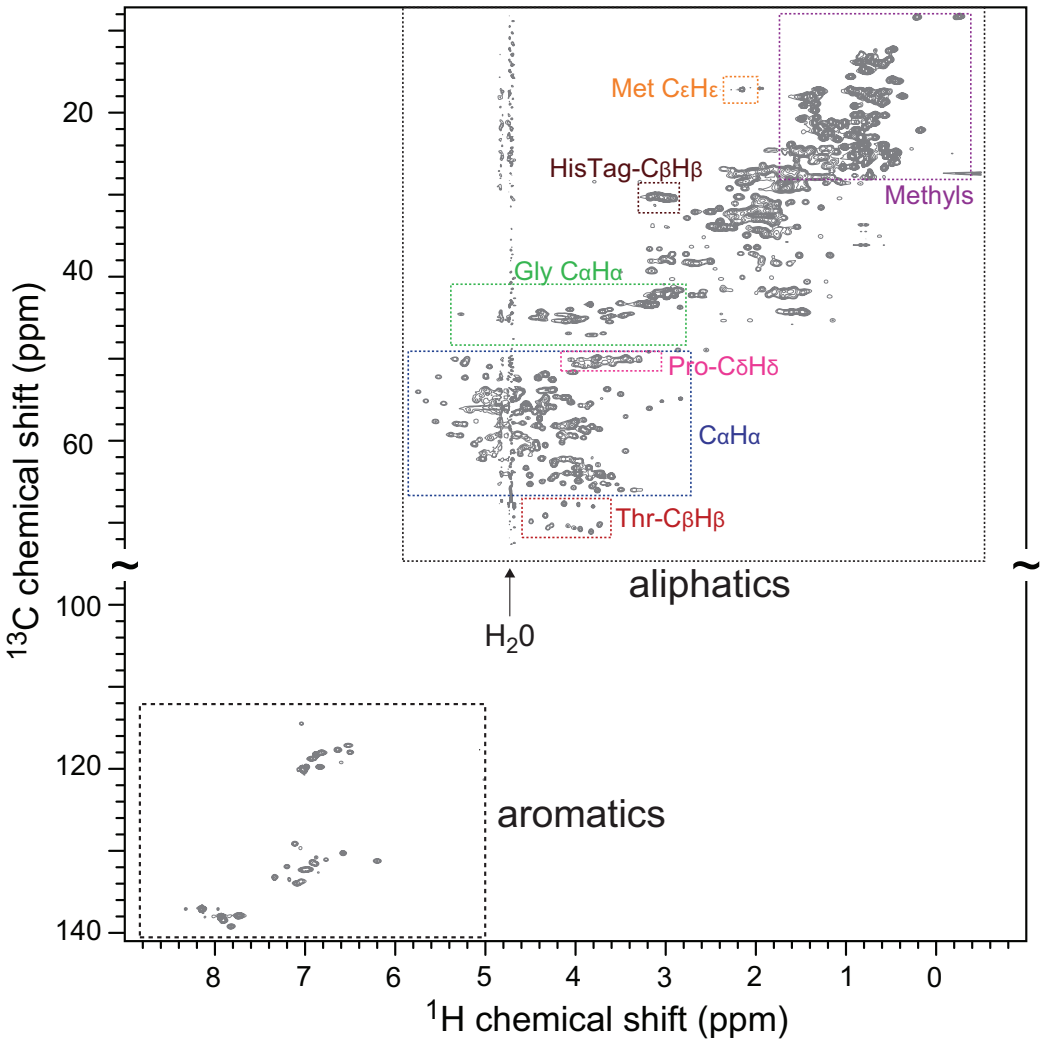


Figure 3.4: ^{1}H-^{13}C correlation protein spectrum of Ldt_{Bs} recorded at 25°C. Some carbon and proton frequencies (highlighted by color frames) are highly specific to residue type and position in the side-chain. Assignment of this 2D spectrum requires the use of 3D spectroscopy.

Aliphatic side-chain

Two strategies are generally used to assign the aliphatic side-chains of proteins. The first one (HCCCONH-TOCSY, (Grzesiek and Bax, 1993)) started with the magnetization on one of the side-chain protons, continued with a transfer to the covalently linked ^{13}C (through $^{1}J_{C-H} \sim 135$ Hz), followed by a transfer to the Cα through TOCSY spin-lock, a transfer to the carbonyl of the same residue (through $^{1}J_{Ca-CO}$), and a transfer to the amide group (through $^{1}J_{CO-N}$), thus connecting the CH groups of a side-chain to the amide group of the following residue in the primary sequence and offering immediate residue specific assignment after the completion of the backbone assignment. The second strategy (HCCH-TOCSY, (Kay *et al.*, 1993)) started similarly but exploited the TOCSY transfer to any carbon in the aliphatic chain to finish by a transfer on the covalently linked proton, thus correlating ^{1}H and ^{13}C nuclei within the side-chain. Run as 3D experiments, each of these strategies imposed a choice for the side-chain nucleus to be edited, ^{1}H or ^{13}C, thus giving rise to a set of 4 experiments (H(CCCO)NH-TOCSY, (H)C(CCO)NH-TOCSY, H(C)CH-TOCSY, and (H)CCH-TOCSY). ^{13}C-edited experiments are usually preferred due to the advantageous dispersion of the carbon signals in comparison to the protons.

The side-chain assignment of Ldt_{Bs} and Ldt_{fm} was mainly performed with ^{15}N- and ^{13}C-edited NOESY-HSQC used to extract distance restraints (see section 3.1.2.1). (H)CCH-TOCSY experiments were used to complete the assignments of the two proteins.

Aromatic side-chain

As aromatic residues can take part of the hydrophobic core in proteins, and are involved in protein active sites as in LDts, the unambiguous assignment of the aromatic ^{1}H, ^{13}C and ^{15}N resonances is essential for an accurate and precise structure determination.

This is nevertheless often a tedious task as the chemical shift dispersion of aromatic nuclei is limited, and as they are separated from the Cβ by a quaternary carbon in position γ which prevent the magnetization transfer in TOCSY experiments (Slupsky *et al.*, 1998). An example of the ^{1}H-^{13}C-correlation spectrum centered on aromatics is shown in figure 3.5 for Ldt_{Bs}.

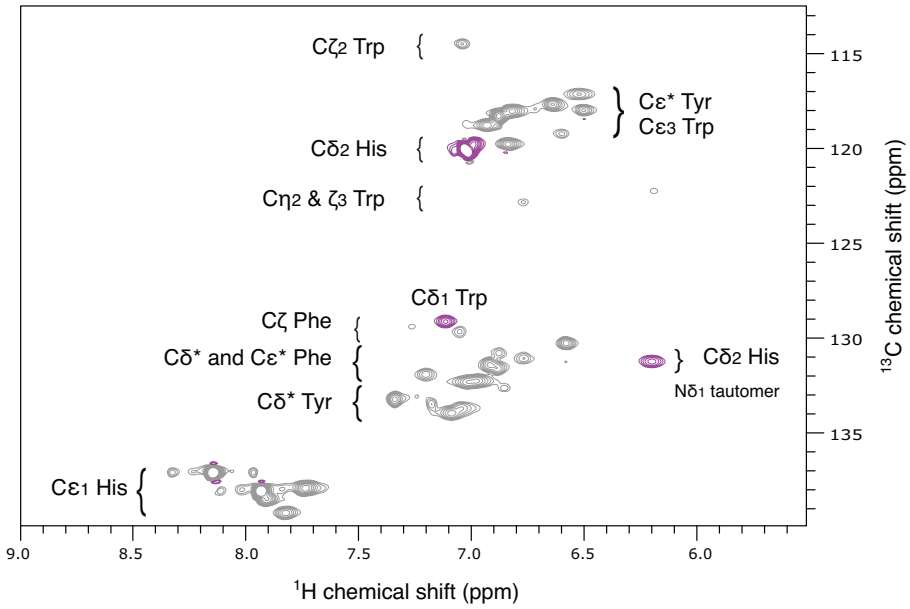


Figure 3.5: Constant-time ^{1}H,^{13}C-HSQC spectrum of the aromatic side-chains of Ldt_{Bs} recorded at 25°C. Correlations are annotated according to typical chemical shift values for aromatic ring nuclei using the IUPAC nomenclature reported in figure 7.23 of the annexes. With a constant-time constant CT = $1/^{1}J_{C-C}$ = 16 ms, correlations from carbons bound to an odd number of aromatic carbons appear negative (in purple) while correlations from carbons bound to an even number of aromatic carbons appear positive (in black). This experiment allows to readily identify His $\delta 2$ and Trp $\delta 1$ correlations. Despite partial signal overlap, detected resonances account for the expected correlations arising from 3 Phe, 7 Tyr, 1 Trp and 10 His (6 of which from the His-tag).

To unambiguously assign these correlations to specific aromatic side-chains in the protein sequence, the group of Lewis E. Kay developed 2D experiments that exclusively transfer the coherence from the Cβ carbon to the Hδ or Hϵ protons within an aromatic side-chain (Yamazaki *et al.*, 1993). These experiments are referred to as (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE, respectively. They were used to complete the assignment of aromatic side-chain (of Phe and Tyr in particular) in LDts and proved to be essential additions to the ^{15}N and ^{13}C-edited 3D NOESY-HSQC despite their intrinsic low sensitivity. In this strategy, the side-chain of His residues remains partial, in part due to the extreme overlap of the C$\delta 2$

and Hδ2 resonances, as shown in figure 3.5. ^{1}H-^{15}N-HMQC exploiting the $^{2}J_{N-H}$ and $^{3}J_{N-H}$ scalar couplings in the imidazole ring provided key assignment information. As use was also made of these experiments to investigate the protonation states of histidines, they are detailed in section 4.2.4.1.

3.1.1.3 Importance of completeness and accurate assignment

Even if the assignment step can resemble an "intermediate" step, to perform as quickly as possible because devoid of direct structural conclusions, it is nevertheless the pedestal of the whole structure determination protocol. The long-considered weakness of NMR in comparison to X-ray crystallography, is that instead of having a direct view of atoms positions through the electronic density maps as in X-ray, the determination of the protein fold by NMR only relies on "indirect" restraints, with a non-negligible part of ambiguous restraints. Therefore, a careful and complete assignment is essential and redundancy is the only way to satisfactorily divert from this weakness. Although the achievement of assignment completeness is not an aim in itself, it is really a prerequisite to a final good structure resolution. In addition, when it comes to the investigation of enzymatic activity, the structural restraints within active sites and/or ligand binding regions must be deeply scrutinized. One of the advantage of NMR with this respect is the possibility of being close to physiological conditions, without need for the addition of artifactual molecule or salt to force crystallization that could disturb the structure of the protein. NMR can therefore provide enzyme structures very close to native structures found in their *in vivo* environment. To fully take benefit of this strength, the assignment step should not be underestimated.

Once both backbone and side-chain assignments are complete and unambiguous, the next step towards structure determination is the extraction of structural restraints, essentially including distances, dihedral and angular restraints.

3.1.2 Extraction of structural restraints from NMR data

Structure calculation by NMR involves an energy minimization step along which all the experimental data available on the macromolecule are taken into account in order to converge towards a unique structure ensemble that satisfies both the hard-coded physical data (such as bond length and torsion angles) and the NMR data. To build the most accurate ensemble, it is therefore necessary to extract a maximum number and a large diversity of structural restraints. The origin and the NMR methods used to extract these data are detailed in the present section.

3.1.2.1 Distance restraints

Unquestionably, the most useful parameter for structure determination is the distance restraint, derived from Nuclear Overhauser Effects (NOEs). The NOE reflects the transfer of longitudinal magnetization between two spins through a spatial dipole-dipole interaction. It depends on the intrinsic magnetic properties of the involved nuclei (gyromagnetic ratio, T_1 and cross-relaxation) and on the internuclear distance. Due to the high gyromagnetic ratio of ^{1}H and to the homogeneous distribution of protons, homonuclear NOEs are usually quantified in proteins. Distances (< 5 Å in non-deuterated protein samples) are extracted from homonuclear 2D or 3D ^{13}C- or ^{15}N-edited Nuclear Overhauser Effect SpectroscopY (NOESY) experiments. As the size of the macromolecule increases, the number of protons is multiplied yielding to drastic correlation overlaps in a 2D ^{1}H-^{1}H-NOESY spectrum (figure 3.6b) and illegibility. Resolution is increased by the addition of a ^{15}N dimension which results in a 3D spectrum

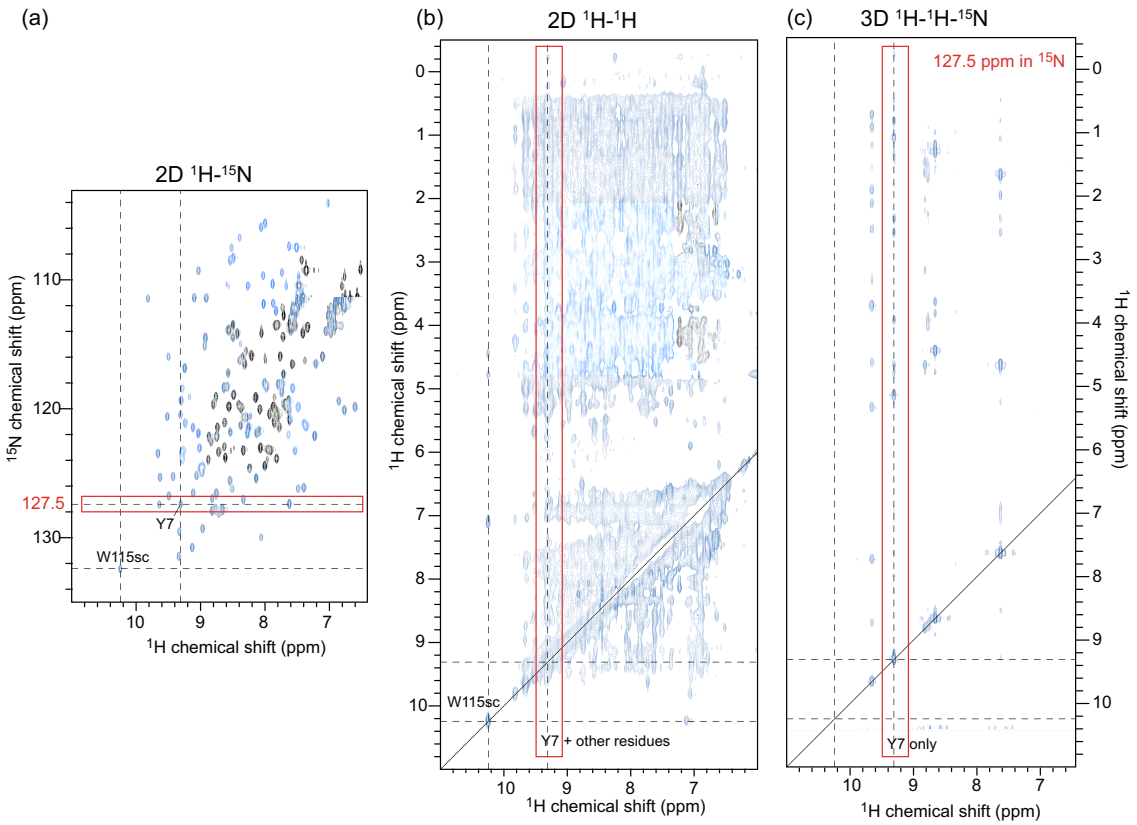


Figure 3.6: Ldt_{Bs} NMR spectra, highlighting the resolution gained from recording 3D NOESY experiments in comparison to 2D homonuclear NOESY experiment. a) 2D ^{1}H-^{15}N-HSQC, b) 2D ^{1}H-^{1}H-NOESY and c) strip of 3D ^{1}H-^{1}H-^{15}N-NOESY-HSQC. While the NOE assignment can be performed on a 2D NOESY experiment for a lonely well-resolved cross-peak, as shown for Trp^{115} side-chain (Nϵ1-Hϵ1), the 2D resolution is not sufficient for most residues to distinguish between overlapped cross-peaks in the proton dimension, as shown for Tyr^{7}. For macromolecules NOE assignment, it is therefore required to separate the peaks in a third dimension (^{15}N or ^{13}C). The strip shown for the 3D-^{15}N-NOESY-HSQC at the ^{15}N-chemical shift of Tyr^{7} (127.5 ppm) demonstrates the quality of the resolution obtained in both ^{1}H and ^{15}N dimensions.

with well-separated cross-peaks (figure 3.6c).

The volume of cross-peaks in NOESY experiments is directly related to the distance separating the two protons:

$$V \propto \frac{1}{r^6} f(\tau_c) \tag{3.1}$$

where V and r represent the volume of the correlation peak, and the internuclear distance, respectively. The effects of global and internal motions of the protein nuclei are taken into account through the function $f(\tau_c)$, where τ_c is the correlation time of the protein. It represents approximately the average time for the molecule to rotate by one radian, and depends on the macromolecular size and on the sample properties (viscosity, temperature ...). This expression is only strictly valid within the Isolated Spin Pair Approximation (ISPA), which never occurs in a protein due to the high proton density. The magnetization thus transfers from one spin to another via other spins in the vicinity. This effect, called spin diffusion, affects the volume of cross-peaks, especially for long mixing times. When choosing the mixing time to collect NOESY experiments, a compromise has thus to be found between the suppression

of spin diffusion (short mixing times) and the time requirement to transfer the magnetization to further spins (long distance restraints in the order of 5 to 6 Å that are crucial to fold the protein). To avoid the potential mistakes in NOE assignment due to spin diffusion artifacts, it is therefore required to search for triangular connections (spin A sees spins B and C: B⟷A⟷C, which are seeing each others: B⟷C), instead of lonely restraints (spin A sees spin B: A⟷B).

To interpret 3D NOESY spectra, which generally contain a huge amount of cross-peaks (several thousands, increasing with the macromolecular size), iterative automatic and semi-automatic approaches developed in the literature have been used for LDts. They will be detailed in the next section (3.1.3).

Before to start a structure calculation, it is often required to provide the software with a restricted set of long distance unambiguous distance restraints in order to favor the convergence of the structure calculation protocol. The Ldt_{fm} and Ldt_{Bs} constructs contain 129 and 175 amino acids, respectively. Among those 32.5 and 40.6 % are methyl-containing amino acids, respectively. As these groups are usually involved in long-distance hydrophobic contacts and present favorable relaxation time constants (long T_1 due to the fast methyl rotation around the C-CH_3 bond), they are particularly good candidates for unambiguous assignment and give rise to key contacts to fold the protein. Inter-methyl NOE correlations must thus be collected but nevertheless be sorted out from other NOESY cross-peaks. The 3D ^{1}H-^{13}C-^{13}C-methyl-selective-NOESY-HSQC designed by Brutscher and coworkers (Van Melckebeke *et al.*, 2004) was recorded for that purpose on each of the apo- and carbapenem-bound forms of LDts with a mixing time of 160 ms. As an example, it offered 170 well resolved cross-peaks manually assigned and quantified in Ldt_{fm} which contains 64 methyl groups.

3.1.2.2 Dihedral restraints

The second type of restraint that is almost systematically used for structure determination is provided by dihedral angles. The peptide backbone of a protein is described by three dihedral angles: ϕ, ψ and ω (figure 3.7 and Table 3.2). ϕ and ψ are defined as the C'$_{i-1}$-N$_i$-Cα_i-C'$_i$ and N$_i$-Cα_i-C'$_i$-N$_{i+1}$ dihedral angles respectively, whereas ω usually describes the trans-configuration of the planar peptide bond (ω = 180°).

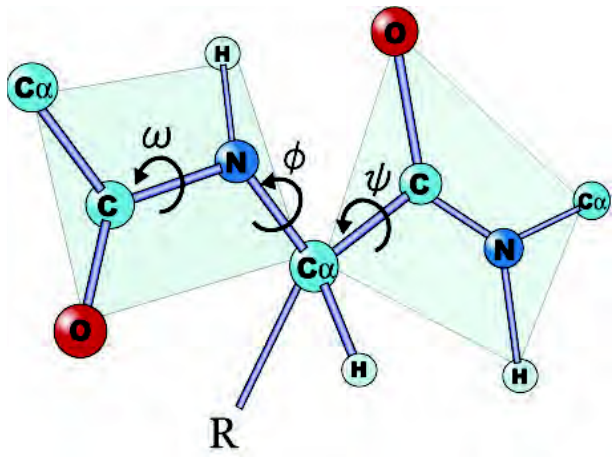


Figure 3.7: Representation of the three dihedral angles describing protein peptide backbone. R group represent the side-chain of amino-acids. The peptide bond is usually in the trans-configuration ($\omega = 180°$).

Secondary structure	ϕ_{min}	ϕ_{max}	ψ_{min}	ψ_{max}
β-sheets	-140	-100	90	130
α-helices	-80	-40	-70	-30

Table 3.2: Range of ϕ and ψ angles classically used for secondary structures in structure calculation (Ramachandran *et al.*, 1963).

Two alternative semi-empirical approaches are used to extract backbone dihedral restraints: one is based on experimental chemical shifts and uses the TALOS+ computer program and the other one is based on the measurement of scalar couplings and their relation to dihedral angles via the Karplus equation.

Determination of scalar coupling constants

Dihedral restraints can also be empirically deduced from ^3J scalar coupling constants. Scalar coupling arise from pure quantum mechanics spin-spin interactions that are mediated by covalent electrons (Fermi contacts).

An empirical relationship between the ^3J scalar coupling constant and the dihedral (or torsion) angle θ was initially proposed by Karplus (Karplus, 1959, 1963):

$$^3\mathrm{J}(\theta) = A\cos^2\theta + B\cos\theta + C \tag{3.2}$$

where A, B and C semi-empirical coefficients depend on the atoms involved in the coupling and the definition of the dihedral angle θ. Karplus coefficients that relate the backbone ϕ dihedral angle to the intra-residue $^3\mathrm{J}_{HN-Ha}$ scalar coupling constant, one of the most commonly measured constant for NMR structure determination, have been determined (Vuister and Bax, 1993):

$$^3\mathrm{J}_{HN-H\alpha}(\theta) = 6.51\cos^2\theta - 1.76\cos\theta + 1.60 \qquad \text{with } \theta = |\phi - 60°| \tag{3.3}$$

The psi dihedral angle can similarly be deduced from the measurements of $^3\mathrm{J}_{Hai-Ni+1}$ heteronuclear scalar couplings (Pellecchia *et al.*, 1998). In contrast to distance restraints, scalar coupling constants only provide local information on the polypeptide chain. Nevertheless, they are particularly useful for an accurate definition of the local conformation, and for the determination of stereospecific assignments (Güntert, 1997). They can also be used for structure validation.

Principle of the TALOS+ program

The TALOS+ program (Shen *et al.*, 2009) utilizes experimental backbone chemical shifts of the protein to derive the ϕ and ψ torsion angles (Cornilescu *et al.*, 1999) through a database containing the NMR and structural data for over 200 proteins. As the chemical shifts strongly depend on local conformation, the program will search within the database for three-three residue fragments with the same sequence that show similar chemical shifts. This artificial neural network approach estimates the likelihood for a given residue to be in a sheet, helix or loop conformation. Residues for which ϕ and ψ angles are predicted with a confidence value larger than 0.7 and for which order-parameters estimated from the chemical shifts are larger than 0.5 are entered in the structure calculation protocol. Today, TALOS+ is widely used by NMR

spectroscopists, as it provides quick and reliable information on secondary structure elements thanks to the predicted dihedral angles (Table 3.2), directly after the chemical shift assignment step and without collection of additional data sets. This approach was used in the case of LDts.

3.1.2.3 Angular restraints

Angular restraints can be derived from Residual Dipolar Couplings (RDCs). This phenomenon occurs if, in the presence of a magnetic field, the molecules in solution exhibit a partial alignment with respect to the imposed field. This alignment results in an incomplete averaging of anisotropic dipolar couplings, yielding to an additional contribution D to the J scalar coupling. RDCs give information on the orientation of individual spin vectors with respect to the magnetic field and thus, for a rigid molecule, with respect to the molecular frame.

If they find their physical origin in the dipolar spin-spin interaction like NOEs, RDCs contain an orientational component in addition to a distance dependence. RDCs are thus complementary to the use of NOEs to provide structural information, and are now widely utilized in NMR structure determinations and refinements (Lipsitz and Tjandra, 2004). RDCs can also be used for *de novo* structure determination but are limited to the backbone structure (Delaglio *et al.*, 2000; Tian *et al.*, 2001; Hus *et al.*, 2001). RDCs are particularly useful in the case of rigid multidomain proteins, as the vectors orientation can be used to define the relative orientation between the domains. In the case of Ldt_{Bs}, the relative orientation between the N-terminal LysM domain and the C-terminal catalytic domain could not be obtained with NOEs due to the scarcity of the interdomain nuclear contacts with distances smaller than 6 Å. Measurement of RDCs and incorporation of the orientational restraints into the structure calculation proved to be a key factor in the obtention of a high resolution structure.

The dipolar coupling between 2 spins I and J (D_{ij}) depends on the internuclear distance between the two spins (r_{ij}), on the orientation of the internuclear vector with respect to the magnetic field (angle θ), and on the molecular reorientation (the $<>$ symbols) (figure 3.8):

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h^2}{8\pi^3 r_{ij}^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \tag{3.4}$$

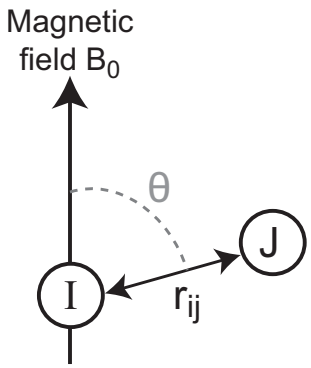


Figure 3.8: Representation of the dipolar coupling between two spins I and J.

In isotropic solutions, the dipolar coupling D_{ij} averages to zero as a result of rotational diffusion. In contrast, in anisotropic conditions, proteins have a preferential orientation yielding to a non zero-dipolar coupling. In weakly aligned media (where approximately one molecule over 10^3-10^4 has a preferential

orientation), the dipolar coupling can be written in the alignment tensor frame (that can be deduced from the molecular frame by an Euler rotation) and simplified to take into account the molecular average orientation, as follows:

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{8\pi^3 r_{ij}^3} \left(A_a (3\cos^2\theta - 1) + \frac{3}{2} A_r \sin^2\theta \cos 2\phi \right) \tag{3.5}$$

where A_a and A_r are the axial and rhombic components of the molecular alignment tensor, and $[\theta, \phi]$ are the polar coordinates of the internuclear vector in the alignment frame. μ_0 is the vacuum permeability, h is the Planck's constant, γ_i and γ_j are the gyromagnetic ratios of nuclei I and J, and r_{ij} is the internuclear distance. A_a and A_r are defined with respect to the diagonal elements A_{xx}, A_{yy} and A_{zz} of the traceless alignment tensor:

$$A_a = \frac{1}{2} A_{zz} \tag{3.6}$$

$$A_r = \frac{1}{3}(A_{xx} - A_{yy}) \tag{3.7}$$

Dipolar couplings are directly related to the alignment of the protein in the magnetic field. Typically,D_{ij} can go from a few Hz in weakly aligned samples in solution (usually always below 100 Hz), up to tens of kHz on fully aligned samples such as solids (Lipsitz and Tjandra, 2004). Some molecules have the natural property to align in a magnetic field due to their anisotropic magnetic susceptibility, such as paramagnetic molecules or phages; this is however not the general case for most macromolecules. Alignment can also be achieved by the use of external components. The most commonly used alignment media for dipolar couplings measurements on protein samples in solution are presented here and include bicelles, bacteriophages, alkyl-poly(ethylene glycol)/alcohol mixture and polyacrylamide gels.

Alignment methods

The first alignment medium used to collect RDCs was the **lipid bicelles** (Tjandra and Bax, 1997). In the appropriate phospholipid mixtures and at specific temperature ranges (Prestegard *et al.*, 2004), the lipids are found in a liquid crystalline phase where they form disk-like structures with 200 to 300 Å diameter and approximately 40 Å thickness (Gaemers and Bax, 2001), often referred to as bicelles. The disks orient with their normal perpendicular to the magnetic field. The most commonly used mixture is constituted of long phospholipid chains of dimyristoyl phosphatidylcholine (DMPC) and of dihexanoyl phosphatidylcholine (DHPC) detergent. Bicelles have been used to measure dipolar couplings in the range of tens of Hz in both proteins and DNA and provide steric alignments.

Then, the use of magnetically aligned **filamentous bacteriophages** was simultaneously introduced in 1998 by Clore *et al.* and Pardi and coworkers (Clore *et al.*, 1998; Hansen *et al.*, 1998). The authors demonstrated that a solution of oriented phage particles forms a liquid crystalline medium, providing partial alignment of macromolecules in solution and yielding to residual dipolar couplings that could be measured in both proteins and DNA (figure 3.9). Advantages of bacteriophages compared to bicelles

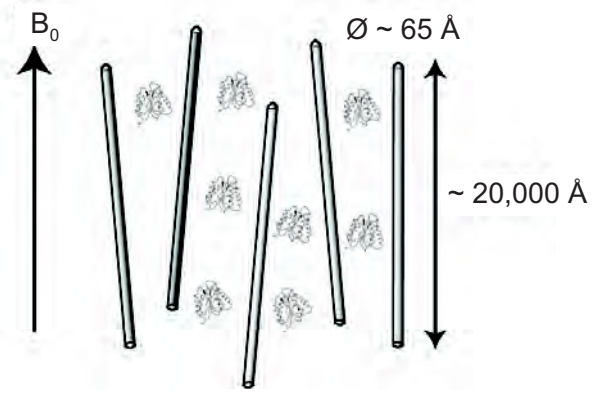


Figure 3.9: Representation of proteins dissolved in Pf1 bacteriophages aligned in the magnetic field B_0. The filamentous phages used for dipolar coupling measurements are approximately up to 20,000 Å in length and 60 Å in diameter. Phages spontaneously align with their long axes parallel to the external magnetic field.

consist in the stability of the alignment medium on a large temperature range (from at least 5 to 60°C) and pH (6.5 to 8.0), and the easily tunable degree of alignment by changing the phage concentration, thereby modulating the size of the dipolar couplings. Furthermore, phages are particularly easy to handle, and the protein can be recovered after centrifugation in the surnatant.

Rückert and Otting introduced another liquid crystalline medium to align proteins and nucleic acids in 2000 (Rückert and Otting, 2000): the **alkyl-poly(ethylene glycol) (PEG)/alcohol mixtures**. It is a lyotropic liquid crystal constituted of adjusted ratios of n(alkyl-PEG) and n-alcohol in water. Alkyl-PEG molecules are denoted as C_mE_n, where m is the number of carbons in the n-alkyl group and n is the number of glycol units in the poly(ethylene glycol) moiety. Phase diagrams have been reported (Freyssingeas *et al.*, 1996) for ternary mixtures of C_mE_n / n-hexanol / H_2O, showing that under certain conditions all the liquid crystalline lamelar phase, referred to as Lα phase, exists only in narrow mixture and temperature conditions. Under the Lα phase, the hydrophobic n-alkyl chains aggregate into planar bilayers, which are responsible for the protein steric alignment when the latter is added to the mixture. The advantages of this method lie in the compatibility of the individual components with proteins, and in its effectiveness over a wide range of temperature (from 10 to 40°C). In addition, this liquid crystalline medium is insensitive to pH and salts, and the constituent chemicals are commercially available and inexpensive compared to the other media available for dipolar couplings measurements. The only practical tricky part in the preparation of this medium stands in the narrowness of the lamellar Lα phase.

Anisotropic aqueous **polyacrylamide gels** (Tycko *et al.*, 2000) can also be used to sterically align protein samples. Anisotropy is induced by straining (compressing or stretching) the gel away from an initial isotropic state within an NMR tube (Chou *et al.*, 2001). The advantages of this method include its efficiency over wide ranges of temperature, pH, and ionic strength and its likely compatibility with a wide range of biopolymers. So as for bacteriophages, the degree of alignment by polyacrylamide gels can be varied by varying the gel density, and/or the gel polymerization and reticulation. The alignment in this medium is independent of the magnetic field direction, instead the orientation is dictated by the direction of the strain.

For Ldt_{Bs}, we chose to measure dipolar couplings in a 11.4 % $C_{12}E_5$ - hexanol mixture with a molar ratio r of 0.96. The protein precipitated in bacteriophages, probably due to attractive electrostatic interactions.

Control of the medium anisotropy

In some alignment media, the degree of alignment can be checked by the measurement of the residual deuterium quadrupolar splitting of D_2O in a simple ^{2}H NMR spectrum. The frequency difference between the two resulting D_2O NMR lines is directly related to the alignment degree of the water imposed by the medium. The alignment of the protein in the medium must be checked independently through the measurement of dipolar couplings (through a ^{1}H-^{15}N-coupled HSQC for example).

Dipolar coupling measurements

The determination of RDCs requires two samples, the isotropic and anisotropic samples. The anisotropic sample only differs from the isotropic sample by the addition of the alignment medium ($C_{12}E_5$/hexanol mixture in the case of LDts) (figure 3.10).

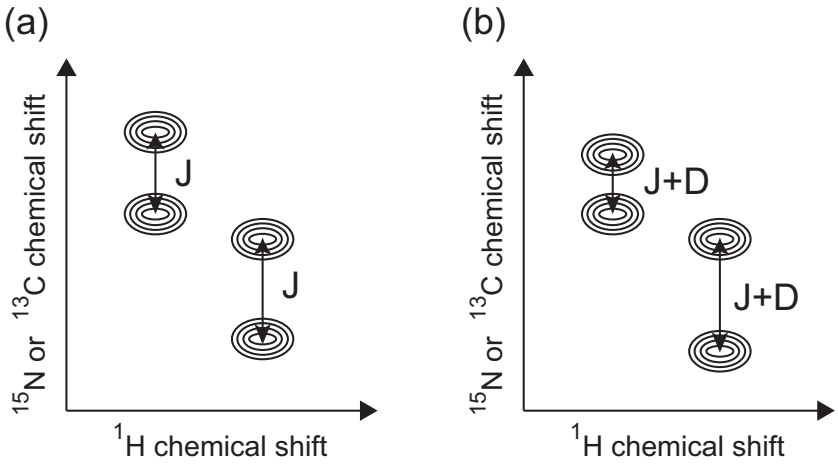


Figure 3.10: Principle of the determination of dipolar couplings. A first experiment with edition of the coupling is performed on the isotropic sample **(a)**. Due to molecular tumbling the dipolar interaction averages to zero and only the J scalar coupling is present. A second experiment is performed in the same conditions on the anisotropic sample **(b)**. A residual dipolar coupling adds to the scalar coupling. The difference between the two measurements offers the dipolar coupling value for each residue.

To measure dipolar couplings, pulse sequences developed for the measurement of J-coupling are used, as scalar and dipolar coupling evolve similarly during pulse sequence delays. Therefore, identical NMR pulse sequences will be collected on both isotropic and anisotropic samples. Due to the edition of the scalar and dipolar couplings, each cross-peak appears as a doublet separated by J or $J + D$ for the isotropic or anisotropic samples, respectively. The difference between the two measurements yields the dipolar coupling D value with its sign. To date, methods for measuring $^1D_{N-H}$, $^1D_{C\alpha-H\alpha}$, $^1D_{C\alpha-C'}$, $^1D_{N-C'}$, $^2D_{H-C'}$ and $^nD_{H-H}$ are available through 2D or 3D experiments (for a review on the pulse sequences used to measure RDCs see (Prestegard *et al.*, 2004)).

As it can be difficult to measure small couplings (especially two-, three- or four-bonds couplings), more evolved methods have been developed, that either increase the value of the coupling by a known scaling factor, or that separate the two components of the doublet in two different spectra. The first method consists in letting the coupling evolve for a longer delay (λt_1) than the one required for the lone chemical shift evolution (t_1). The components of the doublet are thus separated by a factor $\lambda.J$ in the indirect dimension. Furthermore sensitivity is always an issue in these experiments. First the intensity of the initial cross-peak is dispersed in two fractions due to the scalar and dipolar splitting. Second, the accuracy of the determination of the dipolar coupling depends on the signal-to-noise ratio of the experiment. The BEST approach has thus been proposed by the group of B. Brutscher to compensate for the sensitivity loss, yielding to a new set of experiments combining these two implementations (Rasia *et al.*, 2011). A set of BEST-HNCO-J_{NH} and BEST-HN(CO)CA-J_{CH} experiments have been collected on isotropic and anisotropic Ldt_{Bs} samples at 25°C and pH 6.5 to determine the ${}^1D_{N-H}$ and ${}^1D_{C\alpha-H\alpha}$ / ${}^4D_{H\alpha-H_N}$, respectively.

Refinement with RDCs

To use RDCs in a structure calculation, the alignment tensor must be determined, *i. e.* the eigenvalues A_a and A_r as well as the orientation of the alignment tensor with respect to the molecular frame (the Euler angles α, $\beta$and γ). When a preliminary structure is available (from a previous structure calculation performed without RDCs or from a known X-ray structure), the program MODULE can be used to determine the alignment tensor parameters for individual or multiple domains from residual dipolar couplings (Dosset *et al.*, 2001). Experimental D_{ij}^{exp} values and corresponding errors σ_{ij} are given to the program with the preliminary structure. Tensor eigenvalues and eigenvectors are then extracted using least-squares minimization of the target function over all couplings associated with the user-selected domain(s):

$$\chi^2 = \sum_n \left[\frac{(D_{ij}^{exp} - D_{ij}^{calc})^2}{\sigma_{ij}^2} \right] \tag{3.8}$$

where D_{ij}^{calc} are the dipolar couplings back-calculated from the structure using equation 3.5.

The minimization algorithm searches the $(A_a, A_r, \alpha, \beta, \gamma)$ parametric space by random variation of these parameters, using a combination of simulated annealing and minimization protocols (Dosset *et al.*, 2001). Once these parameters are determined, they can be fed into the structure calculation protocol to integrate experimental RDCs as additional restraints. Both the initial global alignment tensor and orientation of the individual internuclear vectors must be optimized. However, in most of the structure calculation softwares, the global tensor can not evolve during the structure calculation process. Consequently, several cycles are often needed between the RDC analysis software like MODULE and the structure calculation protocol like ARIA, until a convergence is reached. In the case of Ldt_{Bs}, the alignment tensors were determined separately for the two domains, and rotated until coincidence of the two alignment tensors, in accordance with experimental data. The two domains can subsequently be considered as one single entity. The difference between Ldt_{Bs} structures obtained in absence and in presence of RDCs restraints is shown in figure 3.11, highlighting the improvement of quality structure brought by the RDCs, as shown by the lower Root Mean Square Deviation (RMSD) values within the 20-lowest

energy structures and a better orientation of the two domains towards each others (by comparison with the X-ray structure).

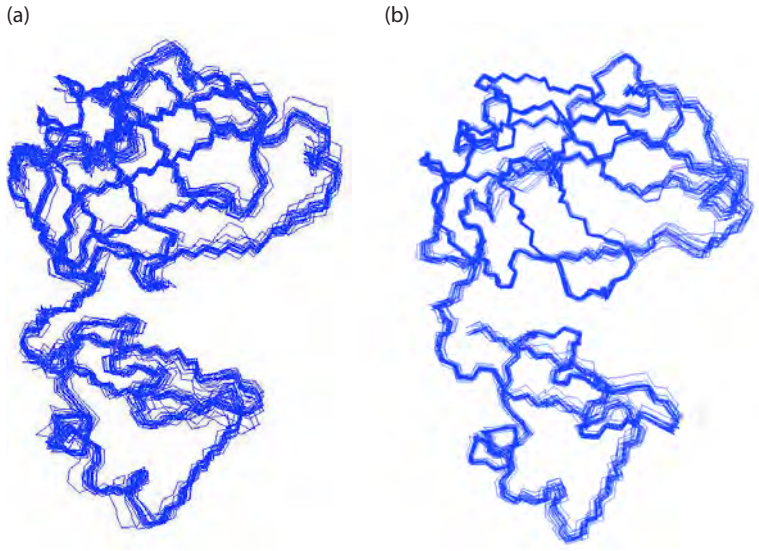


Figure 3.11: Structure of Ldt_{Bs} a) without and b) with RDCs. The use of RDCs resulted in a structure ensemble with a decreased-RMSD from 0.2 unit for both backbone and heavy atoms, and with a relative orientation of the two domains in better agreement with X-ray structure.

All the parameters presented in this section (distance, dihedral and angular restraints) can then be introduced into the structure calculation software in addition to chemical shifts and contribute to build an accurate and realistic protein structure.

3.1.3 Structure calculation and NOE assignment

The determination of a 3D protein structure by NMR consists in finding an ensemble of conformations corresponding to energetic minima, in accordance with experimental data. A randomly folded starting structure is first calculated from the empirical data (bond lengths, bond angles, ...) and the known amino acid sequence. The structure calculation program then tries to fold the starting structure in such a way that the experimental restraints must be satisfied in the calculated structures. To achieve this, an energetic target function is used, that contains a physical term (including empirical parameters, that will depend on the force field used), and an experimental term (including distances and angles). The target function will permit to measure the agreement between a calculated conformation and the structural restraints. This energetic function will be minimized until a global energetic hole is found, yielding to low experimental energetic penalties.

The assignment of NOESY spectra is the most time-consuming step in the analysis of NMR data due to the large number of assignment possibilities, peak overlap and potential artifacts in the spectra. It is usually performed in an iterative fashion with the structure calculation. Several programs are available to perform automatic NOE assignments and high-resolution structure calculations (Altieri and Byrd, 2004) among which: PASD developed by the group of Clore, a program that uses the XPLOR/NIH Molecular Dynamics (MD) engine (Kuszewski *et al.*, 2004); ARIA developed by the group of Nilges, a program that uses the CNS/XPLOR MD engine (Bardiaux *et al.*, 2012); UNIO developed by Güntert

initially (ATNOS/CANDID) and implemented by Herrmann, a program that uses the CYANA/DYANA MD engine (Guerry and Herrmann, 2012). Structure calculation and NOE assignments of LDts (Ldt_{Bs} and Ldt_{fm} in both apo- and acylated forms) were done with the combined use of the UNIO and ARIA softwares.

UNIO (Guerry and Herrmann, 2012) was firstly used to perform the automated peak-picking of 3D ^{15}N- and ^{13}C-edited NOESY-HSQC spectra (not implemented in ARIA), the NOEs assignment, and to find the global fold of the protein. Then, ARIA (Bardiaux *et al.*, 2012), that incorporates a correction for spin diffusion using a relaxation matrix analysis, was used to improve the accuracy and the percentage of the NOE assignments, to include experimental RDCs (in the case of Ldt_{Bs}), and to refine the structures ensemble in explicit solvent.

Both UNIO and ARIA are using simulated annealing during the structure calculation process. These two programs work in the torsion-angle space, where restraints are included as pseudo-harmonic potentials. The simulated annealing protocol starts from random structures, and involves several steps. The initial heating phase reaches very high temperature (until 10,000 K) during which the protein is allowed to fold under the pressure of a force field. During the second phase, the system evolves at high temperature, exploring the conformational space, looking for a potential well. A final gradually cooling phase is applied, during which the system can be stabilized in the potential well previously determined. Some terms of the force field that were inactive during the high temperature steps are used at this stage, such as the non-bonded terms, thus converging onto a more realistic structure.

3.1.3.1 UNIO

UNIO combines ATNOS (which performs the analysis of NOESY raw data), CANDID (which carries the iterative cross-peak assignments and CYANA/CNS (for Crystallography and NMR System (Brünger *et al.*, 1998)) softwares. It is highly efficient in determining a correct fold on the first cycle and performs better than ARIA at that stage.

The experimental input data in UNIO include the primary sequence, the overall chemical shifts list coming from backbone and side-chain assignments, the NOESY processed spectra in Xeasy format without peak-picking, and optional conformational constraints, such as dihedral restraints and/or some manually-assigned NOEs (methyl-methyl NOEs determined for both LDts for example).

In the UNIO protocol (figure 3.12), automated NOESY peak-picking and NOE signal identification is performed with the ATNOS algorithm in association with automated NOE assignment via the CANDID algorithm. For an optimal result, NOESY spectra of high quality are required, as well as accurate calibration of the input chemical shifts to the NOESY spectra. Algorithms used by UNIO for the automated peak-picking are sophisticated, and are thus trustworthy and very efficient. However, NOE cross-peaks resulting from this automated step must always be manually reinspected as signal overlap and spectral distortion can give rise to mistaken, inaccurate or missing cross-peaks.

During automated NOE assignment, an initial assignment list is created first, with one or multiple assignments for each NOESY cross-peak, determined through fitting of chemical shift values within a user-defined tolerance range. Along the iterations, the initial assignment possibilities are ranked depending on the following contributions (Guerry and Herrmann, 2012):

- closeness of fitted chemical shift with the input list
- local baseline correction and evaluation of local noise level amplitudes
- determination of spectrum threshold parameters
- presence of a symmetric peak
- compatibility with covalent polypeptide structure
- network anchoring
- compatibility with intermediate 3D structure (after one iterative step)

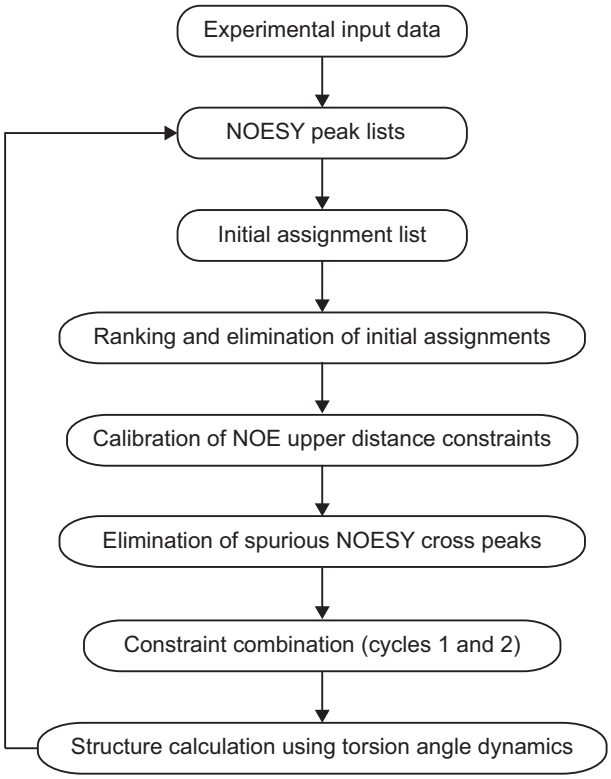


Figure 3.12: Description of the UNIO-based approach used for the initial structure calculation of $\mathbf{Ldt}_{\mathrm{Bs}}$ and $\mathbf{Ldt}_{\mathrm{fm}}$.

The network anchoring evaluates the self-consistency of a specific restraint with the ensemble of NOE assignments, independently from the knowledge of the 3D structure. This approach is based on the research of "triangular" connections, with the identification of NOE cross-peaks with one or several neighboring residues. It is thus a sensitive approach for detecting erroneous and lonely restraints that might artificially constrain some regions of the protein.

Then, peaks are ambiguously or unambiguously assigned and converted into distance restraints with respect to the total peak volume, via the calibration of NOE upper distance constraints. UNIO uses ambiguous distance constraints for handling ambiguities in the initial NOE assignments, a concept introduced in 1995 by Nilges (Nilges, 1995) and which considerably improved the performance of automatic NOE assignment algorithms (Güntert, 2003). Instead of throwing away the NOEs that could not be assigned unambiguously, they are included as ambiguous data into the structure calculation.

Then, spurious NOESY cross peaks are eliminated if their network-anchored score is too low and if they are not compatible with the intermediate 3D protein structure of the preceding cycle.

The step of "constraint combination" in cycles 1 and 2 involves unrelated medium- and long-range

distance restraints which are randomly combined into virtual distance restraints carrying the assignments from two or several of the original constraints. It is a key technique to weaken structural distortions caused by erroneous distance restraints.

Finally, the 3D protein structure is calculated using torsion angle dynamics. The generated structure is used as an input for the following CANDID cycle. The procedure above is repeated several times, typically with 7 cycles, along which NOEs assignment is refined with the help of the structures obtained in the preceding cycle. There are two features that make the UNIO program particularly robust, namely the restraint combination and the network anchoring. However, at the end of the structure calculation, there is often a non-negligible ratio of remaining unassigned peaks. The fact that UNIO is working with a simplified force field, that enables a quick convergence for the protein fold, but can result in an accuracy decline for the final structures ensemble is also to consider. Consequently, we decided to add another step in the structure determination protocol of LDts, with the use of a second structure calculation program (ARIA), which utilizes a more complete force field.

3.1.3.2 ARIA

Similarly to UNIO, the ARIA2.3 (for Ambiguous Restraints for Iterative Assignment) (Bardiaux *et al.*, 2012) program uses an iterative protocol and the concept of ambiguous distance restraints (Rieping *et al.*, 2007).

The ARIA protocol consists in partial NOE assignment, in calibration of NOESY spectra, in merging of distance restraints, and in violation analysis. In the first round of calculations (iteration zero), an initial ensemble of structures is calculated based on the manually prepared assignment list. In each following iteration NOE assignments, calibration and violations analysis are based on the average distances calculated from the lowest energy conformers from the previous iteration (Nilges *et al.*, 1997).

One of the most powerful tool of ARIA resides in its structure-based filters used for the elimination of erroneous assignments at each cycle. However, for the initial iteration (iteration 0), ARIA requires a preliminary fold or an assignment list with abundant long-range distances, otherwise the program might not be able to find a correct initial fold. ARIA is therefore generally not used for a *de novo* structure determination, both because of the difficulties to find the initial fold and because the powerful filters mentioned above are not yet operational at that stage. On the other hand, the algorithm used by ARIA is particularly efficient for improving and completing the NOE assignment once a correct preliminary fold is available (Güntert, 2003). Furthermore, the final step of refinement in explicit solvent (water) is particularly accurate in ARIA/CNS protocol. The PARALLHDG/OPLS hybrid force-field is used, enabling a realistic representation of covalent and non-bonded interactions. a refinement *in vacuo* would lead to artifacts due to missing van der Waals interactions with solvent, which results in unrealistic packing of flexible loops and side-chains. A schematic description of iterative process of ARIA / CNS is shown in figure 3.13.

The use of ARIA subsequent to UNIO is therefore particularly adapted, as it combines the powerful algorithms of UNIO for peak-picking and quick initial fold-determination, with the advantages of ARIA mentioned above.

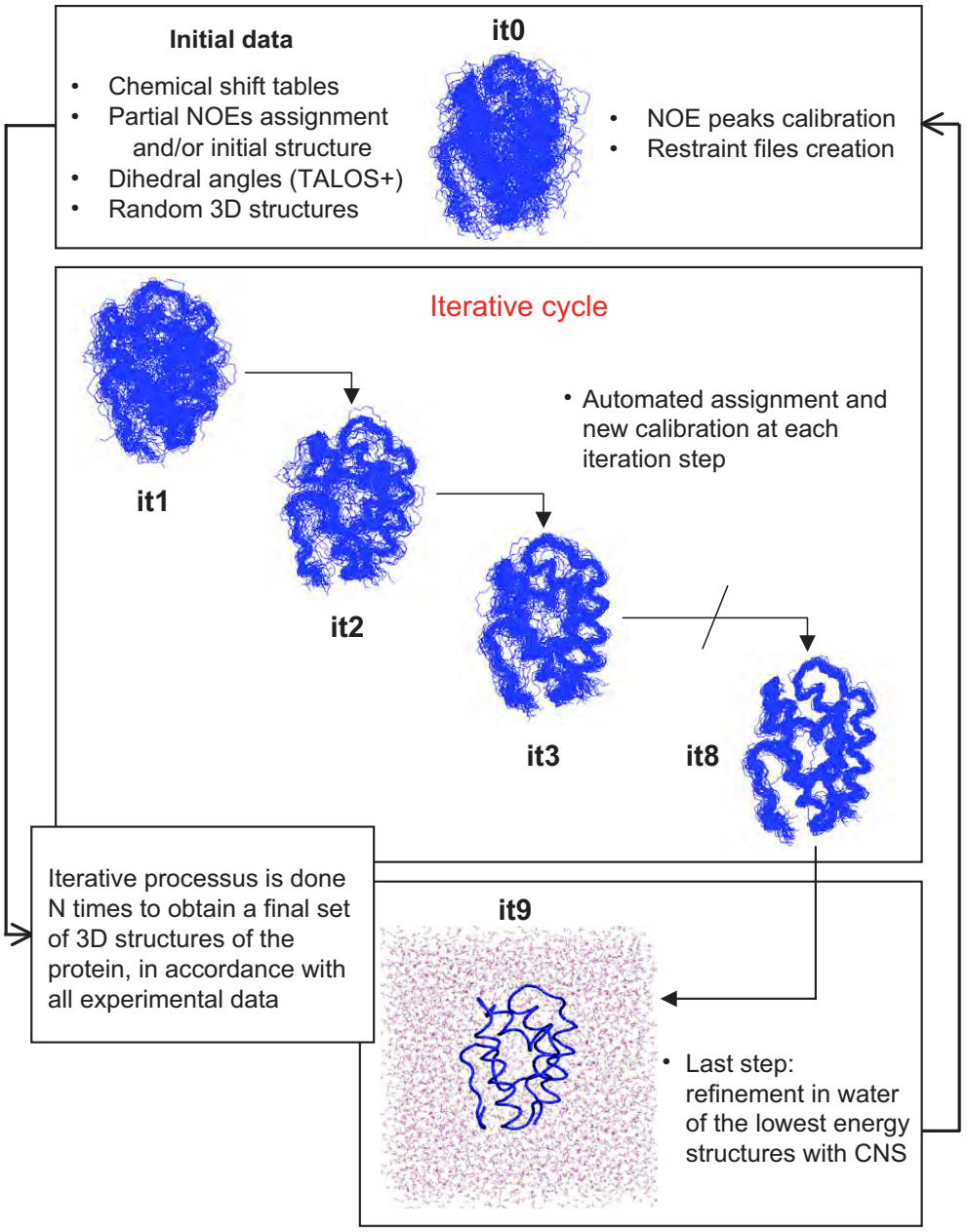


Figure 3.13: Schematic description of iterative process of ARIA / CNS. This process was used for structure calculation of Ldt_{Bs} and Ldt_{fm}.

3.1.4 Structure validation

The determination of NMR protein structures is the result of an elaborate process with many experimental and computational steps. Many structures only rely on NOE contributions, which are nevertheless indirect structural restraints. Consequently, the resulting structure ensembles can contain errors, and sometimes the same input data can give rise to several different folds. It is therefore essential to properly assess the quality of resulting structures, and this step should not be neglected in the whole structure determination process.

In order to detect potential errors, it is important to check the quality and violation reports often generated by the structure calculation program itself. RMSD values, number of violations, Q-factors, and other parameters assess the quality of the structure ensemble by indicating the agreement between the

experimental restraints and the resulting structure ensemble. However, a good RMSD is not necessarily a proof of a good structure quality. On the final ensemble, there should not remain persistent and systematic violations of experimental restraints. Proper covalent geometry of the obtained structural ensemble can also be checked on the basis of their distributions and average values. This includes bond lengths, bond angles, and planarity.

Validation of NMR-derived structures requires dedicated programs that can accommodate the multitude of different experimental data, and that are capable of properly handling multi-model ensembles in their analysis routines (Breukels *et al.*, 2011), such as CING (Doreleijers *et al.*, 2012). For LDts, the Protein Structure Validation Suite (PSVS) program (Bhattacharya *et al.*, 2007) was used to check the quality of the structures obtained. PSVS treats both the structure ensemble and the input data, and interfaces with a number of structure validation programs and routines, including the popular PROCHECK-NMR (Laskowski *et al.*, 1996). Validations are also independently performed by the databases in which NMR data (BMRB) and structural data (PDB) are deposited.

3.2 NMR structure determination of non-covalent and covalent enzyme-inhibitor complexes

One of the main focus of the present work on L,D-transpeptidases is to study the non-covalent and covalent interaction between this class of enzymes and β-lactams, carbapenems in particular. In the case of carbapenems, the formation of the acylenzyme (covalent bond between the active site cysteine of LDts and the carbonyl carbon of the β-lactam ring) is prevented by the use of a Cys-into-Ala mutant of the enzyme. Different NMR methods are available to structurally characterize these complexes. First the antibiotics binding sites can be mapped onto the protein via the chemical shift mapping method. Then, for high-affinity complexes, the structure can be solved if structural restraints (especially intermolecular NOEs) can be obtained between the two molecules. For low-affinity complexes (non-covalent complexes in the case of LDts), the software HADDOCK can be used to create a model for the structure.

3.2.1 Chemical shift perturbation induced by a ligand

^{1}H-^{15}N-HSQC (or ^{1}H-^{15}N-TROSY for proteins of larger sizes) two-dimensional NMR experiments can be used to efficiently monitor binding events involving proteins, as each amino acid except proline gives rise to a single resonance with a position affected by the local chemical environment. The "chemical shift mapping" method consists in quantifying the Chemical Shift Perturbations (CSPs) that are induced on the protein upon addition of the ligand (figure 3.14). Three situations can occur in case of a protein-ligand interaction: (i) the protein resonance shifts progressively as the ligand is titrated in (this situation is called, fast exchange); (ii) the protein resonance tend to disappear (this situation is called intermediate exchange); (iii) the protein resonance intensity decreases at the same time than a new resonance for the same residue appears in another part of the spectrum (this situation is called slow exchange). In the two extreme cases (i) and (iii) the CSP can be determined on a per-residue basis (figure 3.14)). The affected residues can be either directly involved in the interaction, or involved in a global structural change occurring upon ligand binding. Once resonances are assigned (see section 3.1.1), the affected residues can be mapped on the structure of the ligand-free protein and the binding site can be identified. If the changes are spread over the structure a conformational rearrangement is induced by the ligand binding. As such CSPs are a first step into the investigation of ligand binding.

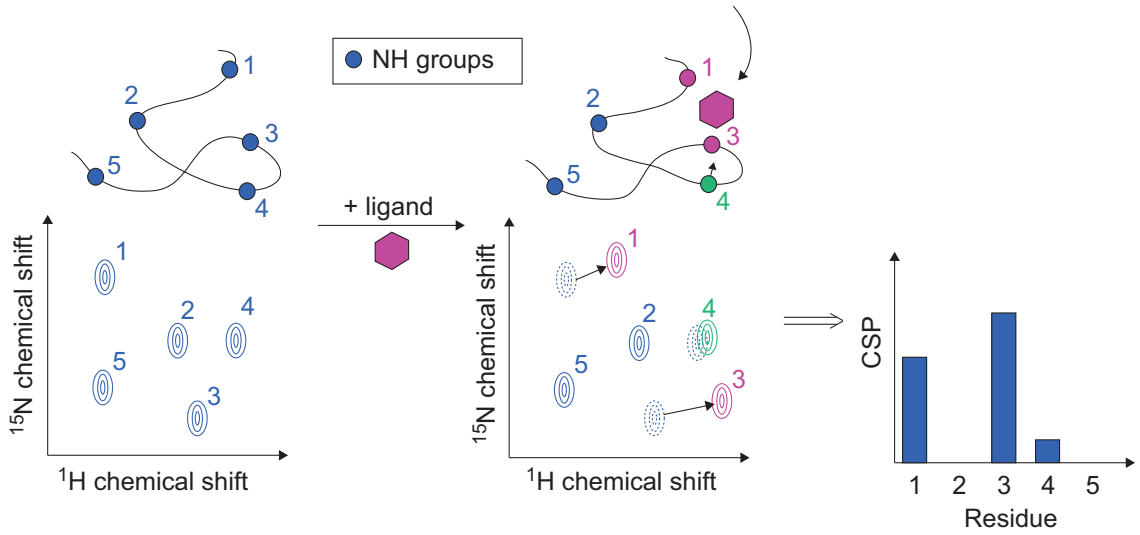


Figure 3.14: Ligand binding monitored by 2D-NMR and chemical shift perturbations in the case of fast exchange. The introduction of a ligand induces perturbation for some residues (here residues 1 and 3, in red). As their environment is modified by the direct proximity with the ligand, their chemical shifts are different, leading to so-called Chemical Shift Perturbations (CSPs). The ligand can also introduce global modifications of the protein conformation, that can disturb residues even if they are not directly in interaction with the ligand (here residue 4, in green). The CSPs can then be reported on a graph in function of the sequence of the protein (right side).

The simplest approach to calculate the CSP uses the normalized length of an Euclidean distance, weighted by the gyromagnetic ratio of nuclei (Schumann *et al.*, 2007):

$$\Delta\delta(\text{ppm}) = \sqrt{(\Delta\delta_H)^2 + \left[\frac{\gamma_N}{\gamma_H}(\Delta\delta_N)\right]^2} \tag{3.9}$$

where the CSP $\Delta\delta$ is the variation of the combined ^{1}H and ^{15}N chemical shifts between the reference spectrum (without ligand) and the spectrum with ligand. $\Delta\delta_H$ and $\Delta\delta_N$ are the variations of chemical shifts in the proton and nitrogen dimensions, respectively. γ_N and γ_H are the gyromagnetic ratio of these two nuclei.

Error on the CSP $d\Delta\delta$ can be estimated as follows (the error is over-estimated in that case):

$$d\Delta\delta(\text{ppm}) = \frac{1}{2\Delta\delta}\left(2\Delta\delta_H\, d\delta_H + 2\left(\frac{\gamma_N}{\gamma_H}\right)^2 \Delta\delta_N\, d\delta_N\right) \tag{3.10}$$

where $d\delta_H$ and $d\delta_N$ are the estimated absolute values of the errors committed on the determination of chemical shifts in the ^{1}H and ^{15}N dimensions, respectively. For LDts, errors were estimated as follows on the basis on the spectral resolution and the signal-to-noise ratio: $d\delta_H$ = 0.005 ppm and $d\delta_N$ = 0.02 ppm.

Compared to other biochemical techniques, NMR offers the possibility to map the interaction on a per-residue basis in a fast, efficient and convenient way, except when K_{Ds} are typically in the order of micro- to milli-molar (intermediate exchange case). The technique is limited by the amount of ^{15}N-labeled protein that is required for the experiment (typically over 10 µM) and the protein size (typically less than 40-50 kDa). When the protein size increases, the detection of ^{1}H-^{15}N correlations can be replaced by the detection of ^{1}H-^{13}C correlations in methyl-groups. These approaches nevertheless require

more elaborate labeling schemes (Ruschak and Kay, 2010). Additional NMR parameters can be followed on the protein or on the ligand by addition of the binding partner: transferred NOE, relaxation (T_1, T_2 and cross-correlated relaxation), diffusion rates, intermolecular magnetization transfer (STD, Waterlogsy...)... (Carlomagno, 2005). All these developed NMR approaches are nevertheless more efficient in the weak affinity regime (typically with 1 μM $< K_D <$ 1 mM).

In favorable cases, the CSP information can be turned into structural restraints (see Docking with Haddock in the following section). In other cases, additional intermolecular restraints need to be collected in order to determine an *ab initio* structure as described in section 3.1.3.

3.2.2 Structure determination of low- and high-affinity complexes

The ultimate goal when looking at protein-ligand interactions is to define the structure of the complex. However, chemical shift mapping may not be sufficient even in the presence of a ligand-free protein structural model. Additional structural restraints may be required, and the choice of the constraint and its measurement will depend on both the affinity of the complex and the extent of the structural rearrangement of the protein upon ligand binding. Structural intermolecular restraints are usually difficult to obtain in cases of weak complexes, due to low quantities of the protein complex in the mixture and a weighted-averaging of the NMR parameters over the different forms of the protein. With K_{Ds} in the order of few tens of mM, non-covalent complexes of LDts with β-lactams fell into this category. On the other hand, covalent complexes of LDts (acylenzymes) were good candidates for the extraction of intermolecular restraints.

3.2.2.1 Low affinity complexes: the docking procedure with HADDOCK

For low affinity complexes, it is often tedious to get intermolecular restraints, thus preventing for a conventional structure calculation protocol. When CSPs suggest that the structural reorganization of the protein is very modest, a rigid body molecular docking program can be used to generate a trustful model.

A large number of docking programs have been developed that are driven by a combination of energetics and shape complementarity between two molecules. The main advantage of using the HADDOCK program (for High Ambiguity Driven protein-protein Docking) (Dominguez *et al.*, 2003; de Vries *et al.*, 2010) resides in its approach driven by experimental data, referred to as "ambiguous interaction restraints" (AIRs). These restraints can be derived from any kind of experimental information available concerning residues that are involved in the intermolecular interaction, such as mutagenesis, mass spectrometry, Chemical Shift Perturbation, RDCs, hydrogen/deuterium exchange, or classical NMR distance restraints (NOEs) when available. The user has to define a threshold to differentiate between "active" and "passive" residues, on the basis of CSPs observed upon complex formation for example and solvent accessibility in the free form of the protein. To run HADDOCK, the PDB files of the two partners and the file containing the AIRs are required. The docking protocol consists of three stages including a rigid-body energy minimization, a semi-flexible refinement in the torsion angle space and a final refinement in explicit solvent. Another advantage of HADDOCK lies in its ability to allow conformational adjustments of the molecules during the formation of the complex, not only in the side chains but also in the backbone. This is the major difference with other docking programs. HADDOCK is thus a powerful tool

to obtain a model for a low-affinity complex.

This program was used to try to model the non-covalent β-lactam C142A-Ldt_{Bs} complex with imipenem. The weak CSPs that could only be measured on the protein were nevertheless insufficient to orient the β-lactam on the exposed protein binding site and generate a convincing model. This failure is due to the very low dissociation constant estimated from titrations.

3.2.2.2 High affinity complexes: measurements of intermolecular restraints

Basically, the structure determination of a high-affinity complex can be performed using standard protocols usually employed for the determination of a ligand-free protein, as described in section 3.1. When CSPs suggest a significant structural reorganization upon ligand binding, all the assignment and structural restraints extraction procedures must nevertheless be performed again on the complex. Furthermore, intermolecular restraints must be added into the structure calculation process in order to get a good resolution on the interaction site(s). The two main sources of intermolecular restraints are provided by intermolecular NOEs and RDCs, and may be more easily obtained than in low-affinity complexes.

Intermolecular NOEs

Extraction of intermolecular NOEs is only applicable when the interaction between the molecules is relatively tight (typically Kd below the micromolar range) (Zuiderweg, 2002). The most powerful method to obtain intermolecular NOEs is the isotope-edited NOE method, which unambiguously discriminates between intra- and inter-molecular NOEs. This requires that the two interacting molecules have different isotopic labeling patterns, as for example an unlabeled ligand and a ^{15}N-^{13}C-labeled protein. In this case, the information on the protein can be obtained by means of $^{13}C/^{15}N$-edited, $^{12}C/^{14}N$-filtered experiments, in which the resonances of the spins belonging to the labeled protein are exclusively observed. Similarly, the resonance assignment and the structural parameters of the ligand can be obtained by $^{13}C/^{15}N$-filtered experiments, in which the resonances of the spins belonging to the unlabeled ligand are selectively observed. Once both protein and ligand are assigned, intermolecular NOE between the two partners can be measured in a $^{13}C/^{15}N$ "half-filtered" NOESY (figure 3.15).

^{13}C-filtered NOESY experiments have been recorded on LDt acylenzymes, imipenem-acylated Ldt_{Bs} and ertapenem-acylated Ldt_{fm} with a 180 ms mixing time. While the Ldt_{Bs} acylenzyme failed to provide intermolecular NOEs due to internal dynamics at the β-lactam covalent binding site, few key intermolecular NOEs were detected between the $CH_3(OH)CH$ substituent of the β-lactam ring of ertapenem and methyl groups near the cysteine active site of Ldt_{fm} (see Part III).

Residual Dipolar Couplings

Structure determination of protein-ligand complexes offers an excellent opportunity for the use of RDC-based methodology. When two interacting molecules are oriented with respect to the magnetic field, both molecules will share the same degree of ordering and the same alignment tensor. RDCs can be measured on the individual interacting partners, yielding in the determination of individual alignment tensors. The two molecules can then be oriented relatively to one another by aligning the two tensors (Jain, 2009). RDCs thus represent powerful long-range structural restraints that can be used to define the

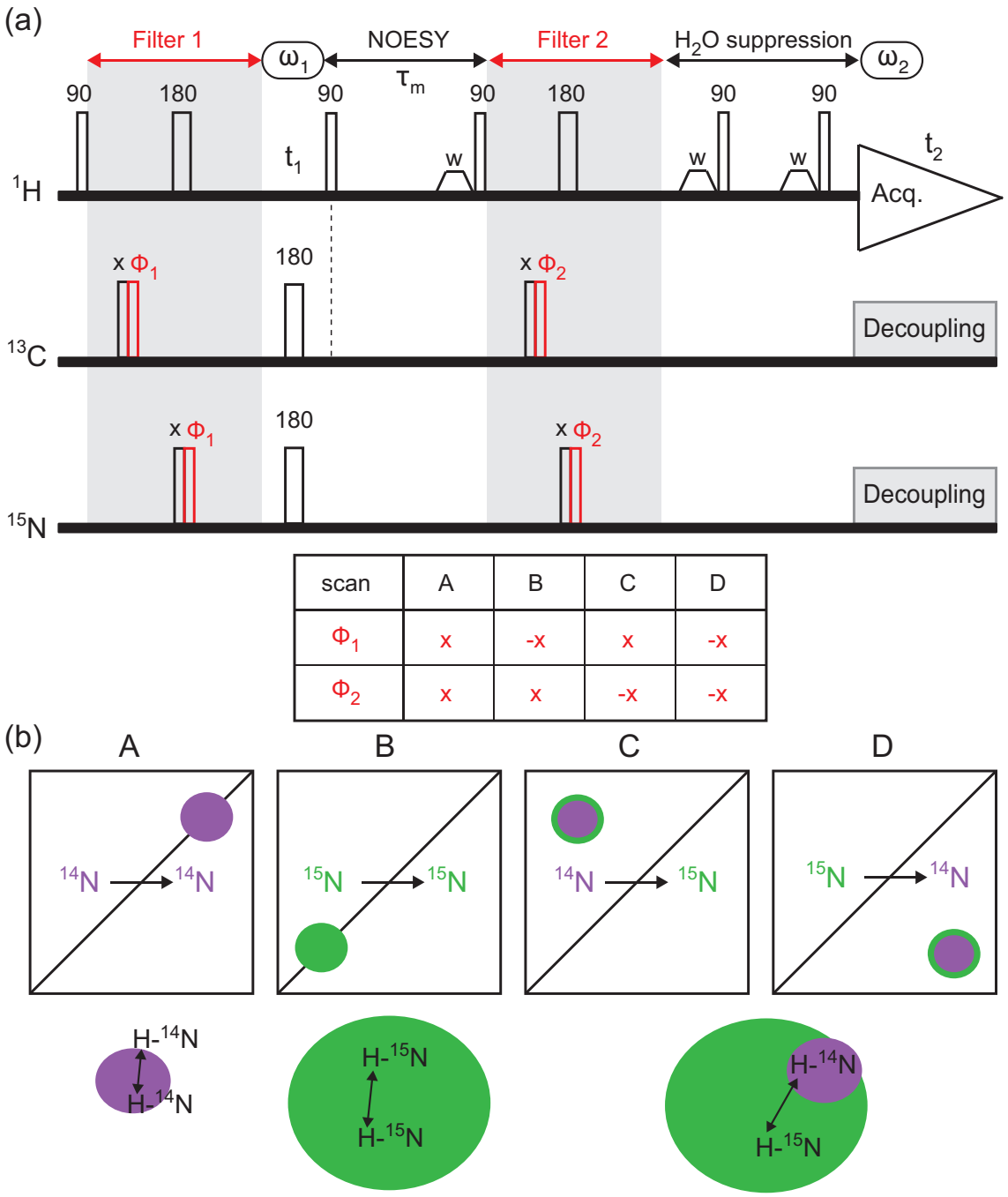


Figure 3.15: Schematic principle of the 2D ^{13}C-^{15}N-NOESY experiment used to extract intra- and intermolecular NOEs between a ^{13}C-^{15}N-labeled protein and its unlabeled inhibitor. a) Simplified NMR sequences of a 2D-filtered NOESY. 4 different experiments can be obtained by changing Φ1 and Φ2 filters. Φ1 permits to select either ^{14}N or ^{15}N-isotopes, and either ^{12}C- or ^{13}C-isotopes before the NOESY, whereas Φ2 applies similar selection after the NOESY. The 4 experiments are collected in an interleaved fashion and sorted out during the processing. Vertical bars represent the radio-frequency90° (narrow) and 180° (large) pulses. "w" represents the water selective pulses, crucial in this experiment for water suppression. τ_M is the mixing time used for the NOE transfer. The phase cycle used for artifact suppression and states-TPPI in the indirect dimension have been omitted. **b)** Schematic resulting spectra of the filtered NOESY. Experiments A and B contain intramolecular NOEs for the unlabeled (ligand) and labeled (protein) partners, respectively. Experiments C and D contain intermolecular NOEs between the two partners: NOEs from the unlabeled ligand (filter 1) to the labeled protein (filter 2), and inversely. The same kind of experiment can be performed by filtering ^{12}C and ^{13}C labeling, either separately or at the same time.

relative orientation of two partners independently of the spatial distance of the probes in the individual partners, as soon as the complex is rigid. However, one should be careful when selecting the alignment medium, as it might cause significant changes in the properties of the sample upon alignment, such as folding topology, interaction between binding partners, or any other functional change (Jain, 2009). Measurement of RDCs on the unlabeled ligand is a difficult task. ${}^{n}D_{H-H}$ RDCs are usually quite weak and difficult to measure in the protein background. ${}^{n}D_{C-H}$ are usually the best candidates but are quite insensitive due to the natural low ^{13}C abundance, unless the protein sample can be obtained at a high concentration (Zhuang *et al.*, 2006). The alternative is a labeling of the ligand by sophisticated organic chemistry. For these later reasons, this approach was discarded for the structure determination of the carbapenem-acylated LDts.

3.2.3 Covalent complexes: implementation of a ligand in ARIA

In the case of LDt acylenzymes, a covalent bond is established between the cysteine sulfur and the β-lactam carbonyl carbon. The structure calculation thus requires to define a modified-amino acid directly within ARIA, to take into account this chemical modification. This is in fact not a trivial task as it requests to modify topology and parameters files within the software protocols. I If the final modified files are given in the Annexes for the interested readers, the principal steps followed to introduce the covalently bound β-lactams on LDts in ARIA 2.3 are briefly introduced here:

1. Creation of the coordinates and topologies of the modified residue. For this purpose, the PRODRG program was used (van Aalten *et al.*, 1996; Schüttelkopf and van Aalten, 2004)
2. Modification of python and xml scripts in ARIA:
   - in the directory src/py/data of the ARIA distribution: atomnames.xml and iupac.xml
   - in the directory src/py/legacy of the ARIA distribution: PseudoAtom.py, Nomenclature.py and AminoAcid.py
3. Modification of CNS files contained in the aria2/cns/toppar directory:
   - topology library: topalldg5.3.pro
   - parameter file paralldg5.3.pro, in case of the introduction of new atoms, new atom types, or new potential terms.

Creation of the ligand coordinates and molecular topologies

Several programs exist to generate coordinates and molecular topologies for small molecules, however they are not all compatible with the various existing packages used during the protein-ligand structural analysis. The program used for LDts was the PRODRG program, as it takes input from existing coordinates or various two-dimensional formats and automatically generates coordinates and molecular topologies suitable for CNS-using softwares (Schüttelkopf and van Aalten, 2004), such as HADDOCK or ARIA.

Three inputs are possible to generate coordinates and/or molecular topologies for a molecule:

- coordinates in PDB format (PDB mode), when available
- description of molecules as a simple ASCII drawing (TXT mode)
- drawing with the popular MDL Molfile/SDfile format (MOL mode), which is used in programs such as ChemDraw and ISIS/Draw

With these three modes, it is possible to modify the chirality of each asymmetric carbon, as well as the protonation states. By default, carboxylates remain deprotonated while guanidinium groups are fully protonated.

Modification of python and xml scripts in ARIA

The following files must be modified within ARIA:

- **AminoAcid.py**: In this file the one-letter and three-letter code associated to the new residue must be added.
- **PseudoAtom.py**: The pseudo-atoms of methylene and methyl groups should be appended in this file
- **Nomenclature.py**: Again, the methylene and methyl groups should be appended in this file, but in a different format compared to the PseudoAtom.py file.
- **iupac.xml**: In this file, every atoms of the modified amino acid are added with their detailed name, type, and corresponding hetero atom (for protons). They are classified as backbone or side-chain atoms, and the N aminyl, N terminus, C amido, C carboxyl, methylenes and methyls are specified.
- **atomnames.xml**: IUPAC, CNS and DYANA names of each atom of the modified residue should be specified in this file.

The detailed modifications of these files are found in Annexe 2. Once they are all correctly modified, the ARIA structure calculation protocol can be initiated, with the automated generation of the sequence.xml file. Yet, to go further it is required to modify the CNS topology and parameters files, that contain all bond distances, angles, dihedrals and impropers, by adding the topology of the ligand.

Modification of CNS topology files in ARIA

Modification of topalldg5.3.pro and paralldg5.3.pro files requires particular attention, as it is not straight-forward and mistakes are easily made. The bonds and angles that are similar to existing amino acids can be defined with the same atom types. This approach simplifies the modification protocol, as all bond distances, angles, dihedrals and impropers values are thus already determined. It can concern for example the backbone of the cysteine in the case of the carbapenem-modified cysteine in the particular case of LDt acylenzymes. It also concerns all aliphatic protons, and moieties often found in protein residues (carbonyls, aliphatic chains, ...). However, for groups of atoms that are not found in existing amino acids, novel types must be added into the topallhdg5.3.pro.

The details of the modifications of these files for the carbapenem-modified cysteine are given in Annexe 2 (Annexes 2.6 and 2.7). Careful implementation of these modifications in ARIA 2.3 enabled the successful structure calculation of the imipenem-acylated $\mathrm{Ldt_{Bs}}$ and ertapenem-acylated $\mathrm{Ldt_{fm}}$.

The quality of the structure obtained for a non-covalent or a covalent complex is conditioned by the determination of intermolecular restraints. Their measurement is nevertheless conditioned by the thermodynamics and kinetics of the complex formation. Such data are worse obtaining before the structure calculation.

Chapter 4

Thermodynamical approach by NMR

NMR is particularly suitable for the thermodynamics of low-affinity complexes ($K_D >$ 10-100 µM), while other biophysical are often limited to tight affinities. The thermodynamics of LDt-non-covalent complexes with different classes of antibiotics was therefore performed using NMR. Furthermore, taking advantage of the atomic resolution of NMR, efficient methods can be set up to extract the pKa of ionizable residues. This is of particular importance for residues in the active site in order to elucidate mechanistic issues of the enzymatic catalysis.

This chapter deals with these two thermodynamical aspects in which NMR is a determinant tool. The method used to determine the dissociation constant of non-covalent LDt-β-lactam complexes is presented first. In the second part the method developed to determine pKa is detailed, with particular emphasis on histidine residues.

4.1 Determination of a K_D dissociation constant by NMR

4.1.1 NMR signal and time regimen a complex formation

Characterization of protein - ligand interaction requires the consideration of both thermodynamical and kinetics parameters. Non-covalent interaction of two compounds induces enthalpy $\Delta_r H$, entropy $\Delta_r S$ and free enthalpy $\Delta_r G$ modifications, according to:

$$\Delta_r G^0 = \Delta_r H^0 - T\Delta_r S^0 \tag{4.1}$$

with $\Delta_r G^0$ and $\Delta_r H^0$ in $kJ.mol^{-1}$, T in Kelvin and $\Delta_r S^0$ in $kJ.mol^{-1}.K^{-1}$.

In the case of a 1:1 protein-ligand complex, where P is the protein (receptor), L the ligand and PL the complex, the reaction is described as:

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL \tag{4.2}$$

Kinetics aspects of interaction are quantified thanks to the association constant k_{on} in $mol^{-1}.L.s^{-1}$ and the dissociation constant k_{off} in s^{-1}.

The thermodynamics of this equilibrium is characterized by the affinity (or association) constant K or the dissociation constant K_D:

$$K = \frac{(PL)}{(P)(L)} \tag{4.3}$$

where (P), (L) and (PL) are the activities at equilibrium of the free form of the protein, the free ligand and the bound form of the protein, respectively. The activity of a molecular solute (X) in a diluted solution is defined as: (X) = [X]/C° with C° = 1 $mol.L^{-1}$.

Thus, one can define a pseudo-constant K' = K/C_0 = [PL]/[P][L] expressed in $mol^{-1}.L$. By analogy, a pseudo-constant can be defined for the dissociation (inverse) equilibrium. This parameter, called the dissociation constant in biology and expressed in $mol.L^{-1}$, is defined as:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}} \tag{4.4}$$

A weak K_D (nM to µM) means a strong affinity, whereas a strong K_D (100 µM to 100 mM) is indicative of a low affinity complex. The lower the K_D, the lower is the required concentration of the ligand in order to form the complex and the better is the potency of the inhibitor in drug-binding assays.

During a titration, the majority species will evolve from P to PL. It is important to note that the system can evolve through three distinct chemical exchange: slow, intermediate or fast exchange. The exchange rates between free and complexed forms can be accessible at the time-scale of an NMR measurement. Generally, at the intermediate exchange rate, the NMR signal is not observable anymore as it becomes indistinguishable from the baseline noise. This occurs when the exchange kinetic constant $k_{ex} = k_{off} + k_{on}[L]$ is in the same order of magnitude than the chemical shift difference between the resonances of the free and bound form of the protein.

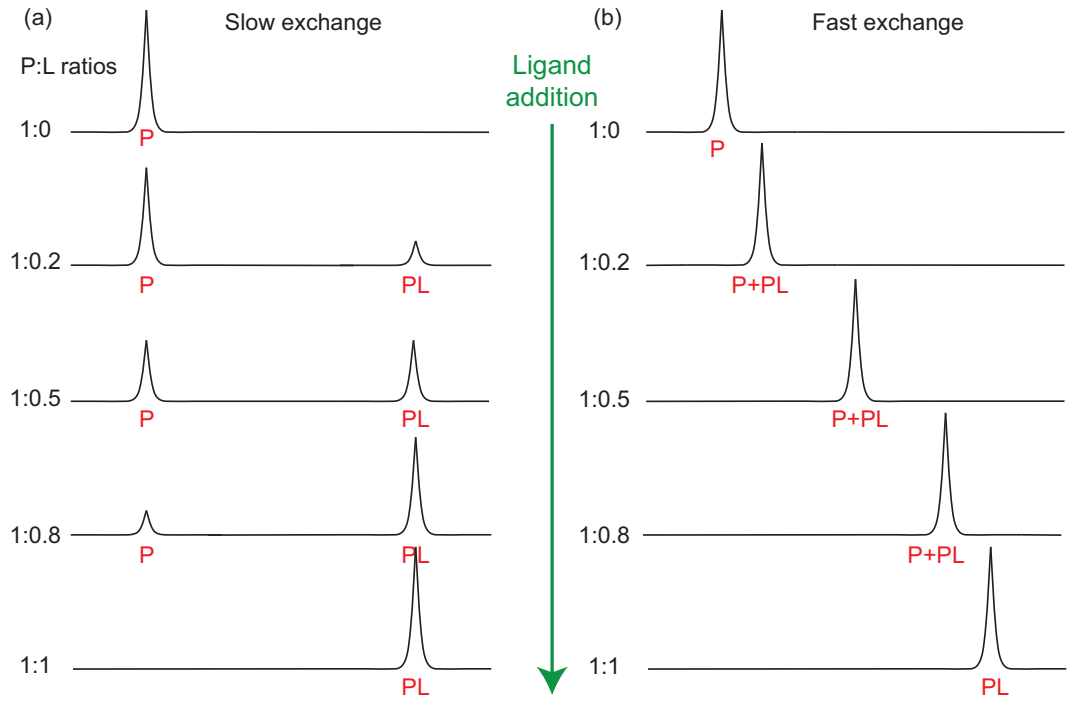


Figure 4.1: Relative position protein NMR signals upon ligand binding in the case of a 1:1 complex - impact of the exchange rate. In the case of slow exchange, two resonances corresponding to the free and bound states are visible along the complex formation, whereas in the case of fast exchange, one resonance corresponding to a weighted average chemical shift between the free and bound states is visible. In the latter scenario, the resonance is moving along the successive additions of the ligand. The protein:ligand (P:L) ratios are indicated on the left.

For a system in the slow-exchange situation, separate resonances are observable for both the free and bound states at their respective chemical shifts (figure 4.1a). Successive addition of ligand quantities to a

protein solution results in the increase of the intensity of the new signal corresponding to the bound protein, and in a concurrent decrease of the intensity of the free protein resonances, reflecting the decreased proportion of free protein during the titration. An integration of these peaks can be done, resulting in the quantitative contribution of both P and PL states along the complex formation. Slow exchange situation is ideal in NMR as it gives the full power of multidimensional NMR methods to characterize the structure and dynamics of the complex (Wand and Englander, 1996).

On the contrary, for a system in fast exchange, the resulting spectra will contain only one resonance corresponding to the weighted average of the frequencies from the free and the bound states (figure 4.1b). Structure and dynamics of the complex are much more difficult to extract, especially in the absence of thermodynamic and kinetic data.

4.1.2 NMR titrations and K_D determination

Along a titration, the protein and ligand concentrations evolve between a free and a bound form, so that:

$$\begin{aligned} [P]_0 &= [P] + [PL] \\ [L]_0 &= [L] + [PL] \end{aligned} \tag{4.5}$$

where $[P]_0$ and $[L]_0$ are the total concentrations of protein and ligand, respectively. [P], [L], and [PL] have the same significance than in equation 4.4, which remains valid. These total concentrations are usually known, because related to the quantity of material introduced in the NMR tube along the titration. The key to the K_D determination is to relate these known solution $[P]_0$ and $[L]_0$ concentrations to the equilibrium concentrations $[P]$ and $[L]$, by means of the NMR observables (Fielding, 2007).

When combining equations 4.4 and 4.5, the second order polynomial following equation results:

$$[PL]^2 - ([P]_0 + [L]_0 + K_D)[PL] + [L]_0[P]_0 = 0 \tag{4.6}$$

The analytical solution to this equation gives the fraction of bound protein $[PL]/[P]_0$:

$$\frac{[PL]}{[P]_0} = \frac{1}{2}\left[1 + \frac{[L]_0}{[P]_0} + \frac{K_D}{[P]_0} - \sqrt{\left(1 + \frac{[L]_0}{[P]_0} + \frac{K_D}{[P]_0}\right)^2 - 4\frac{[L]_0}{[P]_0}}\right] \tag{4.7}$$

Given a protein concentration $[P]_0$ = 100 µM, the graph in figure 4.2 reports the fraction of bound protein $[PL]/[P]_0$ as a function of ligand-to-protein ratio for different values of K_D. It clearly appears from this graph that $[P]_0/100 < K_D < 100[P]_0$ can be differentiated (*i. e.* for the chosen protein concentration: $10^{-6}M < K_D < 10^{-2}M$). Thus to determine a reliable K_D value, the protein concentration must be adapted optimally to the range $K_D/10 < [P]_0 < 10K_D$ (and tolerably to the range $K_D/100 < [P]_0 < 100K_D$). Due to the sensitivity of the NMR experiments, K_D values can be determined from 0.1 µM (with protein concentrations of 10 µM) to a few 100 mM. If one considers that the k_{on} is limited by translational diffusion of the protein and ligand particles and that K_D is thus essentially determined by k_{off}, the exchange regime that is more favorable for the dissociation constant determination is the fast exchange regime (some exceptions are known and a mixture of different regimes can be observed within the same titration without preventing the thermodynamics characterization of the

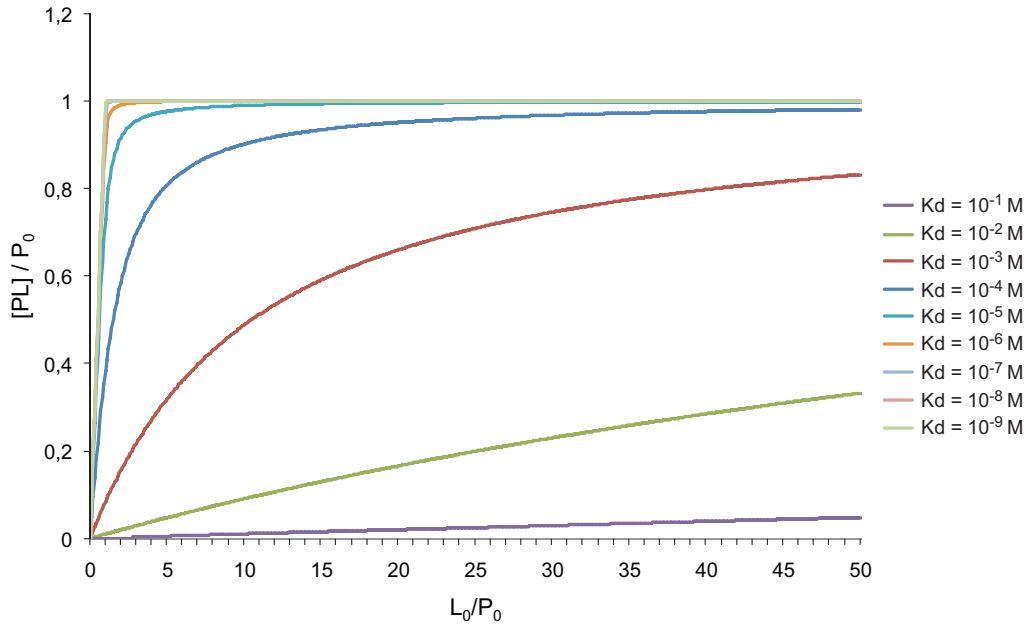


Figure 4.2: Theoretical curves showing the dependence of the bound protein fraction as a function of ligand-to-protein ratio for different dissociation constants. $[P]_0$ is set up to 100 µM.

equilibrium). This regime is thus exclusively considered in what follows.

Chemical shift changes of the protein backbone amide resonances are monitored as a function of ligand concentration along the titration using 2D ^{1}H-^{15}N-HSQC experiments (or TROSY versions for larger protein sizes) (figure 4.3). Chemical shift perturbations, calculated as specified in section 3.2.1 on page 100, can then be reported as a function of ligand-to-protein ratio.

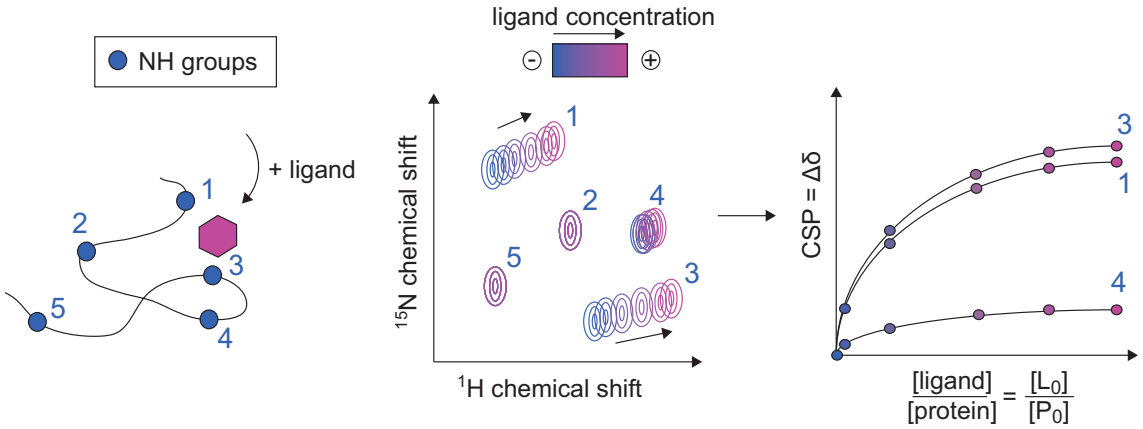


Figure 4.3: Principles of the determination of K_D by NMR titration in the case of fast exchange. The increasing ligand concentration induces CSPs that can be plotted as a function of ligand-to-protein ratio $[L]_0/[P]_0$. Curve fitting through equation 4.9 then provides for the dissociation constant K_D and for the chemical shift difference between the bound and the free forms of the protein.

In the fast exchange regime, the observed chemical shift δ_{obs} is the weighted average of the chemical shift of the free form and the bound form of the protein: $\delta_{obs} = ([PL]/[P]_0)\delta_{PL} + ([P]/[P]_0)\delta_P$. As the percentage of free and bound forms of the protein are linked, $\delta_{obs} - \delta_P = ([PL]/[P]_0)(\delta_{PL} - \delta_P)$,

the CSP $(\Delta\delta) = \delta_{obs} - \delta_P$ is thus:

$$\Delta\delta = \Delta\delta_{max}\frac{[PL]}{[P_0]} \tag{4.8}$$

where $\Delta\delta_{max}$ is the chemical shift difference between the free and bound form of the protein.

After including equation 4.7, the experimental data described in figure 4.3 should be fitted with the following equation to determine K_D and $\Delta\delta_{max}$, from which the chemical shift of the protein complex can be deduced:

$$\Delta\delta = \frac{\Delta\delta_{max}}{2}\left[1 + \frac{[L]_0}{[P]_0} + \frac{K_D}{[P]_0} - \sqrt{\left(1 + \frac{[L]_0}{[P]_0} + \frac{K_D}{[P]_0}\right)^2 - 4\frac{[L]_0}{[P]_0}}\right] \tag{4.9}$$

In situations where $K_D > [P]_0$, the saturation may be far from reached (see $K_D = 10^{-3}M$ in figure 4.2), it is thus more accurate to define $\Delta\delta_{max}$ as a floatting parameter during the fitting process. Errors on $\Delta\delta$ are evaluated as in section 4.3 (see equation 3.10 on page 101). They are used in the Monte Carlo non-linear square fit to estimate the errors on the determined K_D values.

When the complex affinity is weak ($K_D > 1$ mM), high ligand-to-protein ratios must be investigated. The total protein concentration $[P]_0$ thus does not remain constant along the titration. New variables have then to be defined as follows, to preserve a two-parameter fitting procedure:

$[P]_{0i}$ = Initial protein concentration (known and constant)
v_{iP0} = Initial protein volume (known and constant)
$[L]_{0i}$ = Concentration of ligand stock solution (known and constant)
v_L = Added-volume of ligand (known and variable)

$$\frac{L_0}{P_0} = \frac{[L]_{0i}v_L}{[P]_{0i}v_{iP0}} = X \tag{4.10}$$

The dissociation constant can then be fitted with the following equation:

$$\Delta\delta = \frac{\Delta\delta_{max}}{2}\left[1 + X + K_D\left(\frac{X}{[L]_{0i}} + \frac{1}{[P]_{0i}}\right) - \sqrt{\left(1 + X + K_D\left(\frac{X}{[L]_{0i}} + \frac{1}{[P]_{0i}}\right)\right)^2 - 4X}\right] \tag{4.11}$$

For the study of the non-covalent interactions between LDts and β-lactams, equations 4.11 and 4.9 were used to fit the experimental titration data of Ldt_{Bs} and Ldt_{fm}, respectively. The observed chemical shifts corresponding to resonances that are mostly affected by the ligand binding (with CSP cutoff of 0.03 ppm) are averaged for each ligand-to-protein ratio and a single chemical shift perturbation curve is fitted. When different regions of the protein showed different behavior, as it was the case for Ldt_{fm}, chemical shifts were fitted individually with a global dissociation constant using a modified script within ModelXY module of NMRPipe (Delaglio *et al.*, 1995). Alternative 1:2 protein:ligand complexation models are also under development for the analysis of the Ldt_{fm} data.

4.2 Determination of pKa residues in proteins by NMR

Electrostatic interactions are crucial in biology, and proton dissociation constants (pKas) of individual groups in proteins provide sensitive indicators of intra- and intermolecular interactions (Nakamura, 1996). This section deals with general methods used for accurate pKa determination by NMR, and details the particular case of histidines, which requires to take into account the tautomeric forms of the imidazole ring.

4.2.1 Ionizable residues

The presence of charges on protein surfaces is necessary to maintain protein solubility and to prevent aggregation. Positive charges in proteins are provided by the side-chains of His, Arg, and Lys residues, while negative charges are provided by the side-chains of Asp, Cys, Glu Tyr residues.. The seven ionizable residues in proteins (His, Lys, Arg, Tyr, Asp, Glu and Cys) account in average for approximately one third of all residues in proteins (Table 4.1), and are involved in many essential properties of proteins, including structure, function, stability, solubility, dynamics and inter-molecular interactions (Nielsen *et al.*, 2011).

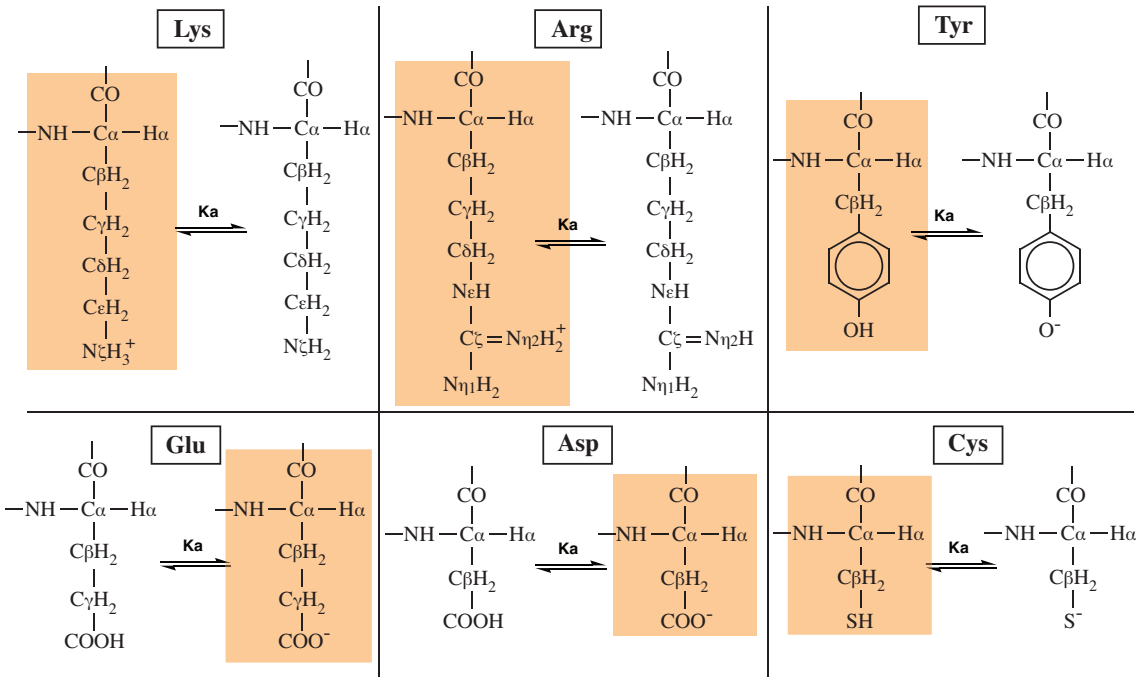


Figure 4.4: Ionizable protein residues. At physiological pH, ionizable residues are generally found in their charged form (Kim *et al.*, 2005), except Tyr and Cys (in orange). Histidines are not represented on this figure as they will be depicted in figure 4.8b.

Charged and uncharged states of ionizable residues are shown in figure 4.4, and their pKa values in Table 4.1. For Asp, Glu, and Cys, the ionizable groups are uncharged below their pKa and negatively charged above their pKa, whereas the ionizable groups of His, Lys, and Arg are positively charged below their pK and uncharged above their pKa. Usually pKa values are reported for free amino acids or for amino acids in short peptides. However, in proteins, these pKa values can drastically differ from pKa_{int} values reported in Table 4.1, especially if the involved residues are implicated in an hydrogen bond network, in electrostatic interactions or are solvent inaccessible.

Group	Content (%)	pKa_{int}	Average pKa	Low pKa	High pKa
His	2.2	6.5	6.6 ± 1.0	2.4	9.2
Lys	5.9	10.4	10.5 ± 1.1	5.7	12.1
Arg	5.1	12.3			
Tyr	3.2	9.8	10.3 ± 1.2	6.1	12.1
Asp	5.2	3.9	3.5 ± 1.2	0.5	9.2
Glu	6.5	4.3	4.2 ± 0.9	2.1	8.8
Cys	1.2	8.6	6.8 ± 2.7	2.5	11.1

Table 4.1: Characteristics of ionizable side chains in proteins. Values are taken from (Pace *et al.*, 2009) and (Grimsley *et al.*, 2009). The pKa_{int} values were calculated for the ionizable side-chains of residue "X" present in blocked pentapeptides with the structure Ala-Ala-X-Ala-Ala while average, low, and high pKa values come from protein experimental data.

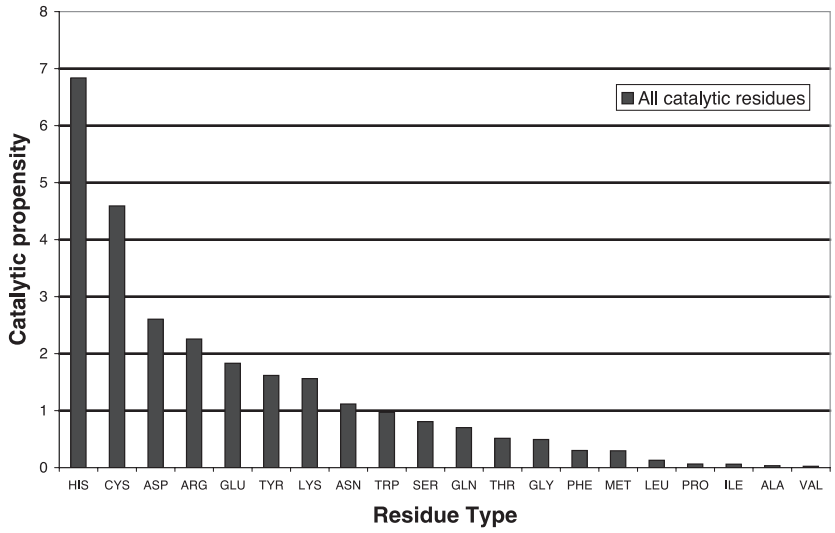


Figure 4.5: Catalytic propensity of amino acids in proteins, taken from (Bartlett *et al.*, 2002). This diagram shows that ionizable residues are highly involved in catalytic steps compared to other protein residues. Histidines are particularly suitable for carrying out these catalytic mechanisms, as their pKa is close to physiological pH and thus an play the role of a nucleophile, an acid, a base, or be involved in the formation of reaction intermediates.

Figure 4.5 demonstrates that the seven ionizable residues show the highest propensity to be involved in catalytic processes, most specifically histidines with the highest catalytic propensity, followed by cysteines, which constitutes 5.6 % of catalytic residues (Bartlett *et al.*, 2002) despite its overall low abundance (1.2 %). In LDts, a Cys and a His were identified in the catalytic site. To fully understand the reaction mechanism of this class of enzyme, it is of significant importance to investigate the protonation states of these two residues in particular.

4.2.2 pH titration

The acid-base equilibrium between the protonated (A) and the unprotonated (B) form of ionizable residues with thermodynamic constant Ka is:

$$A + H_2O \rightleftarrows B + H_3O^+ \tag{4.12}$$

NMR is a suitable technique to investigate amino acids pKa values in proteins as chemical shifts

are very sensitive probes to investigate changes in the environment. This is particular true upon ionization changes, as they will lead to significant variations in the heteronuclear electronic distribution upon (de)protonation. The method for accurate pKa determination involves a pH titration, with the widest pH-range possible depending on the bearable limits for the protein of interest. The chemical shifts of several nuclei for each ionizable residue of interest are thus monitored as a function of pH. A representation of ideal curve that can be obtained during a pH titration is shown in figure 4.6. The observed chemical shift δ_{obs} is the weighted average of the chemical shifts of the protonated δ_A and unprotonated δ_B, ponderated forms:

$$\delta_{obs} = f_A\delta_A + f_B\delta_B \tag{4.13}$$

where f_A and f_B are the fractions of protonated and unprotonated forms, respectively.

Therefore, in acidic environment, δ_{obs} will be representative of δ_A, whereas in basic environment, δ_{obs} will be representative of δ_B (figure 4.6). The acido-basic reaction always corresponds to a fast-exchange, compared to the NMR time-scale. The chemical shifts can therefore be easily followed as the titration progresses.

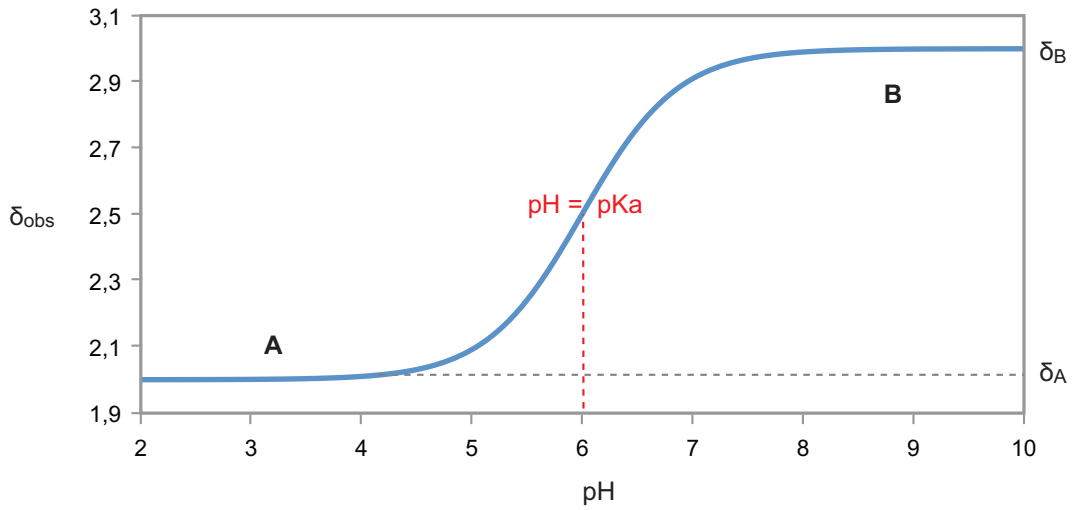


Figure 4.6: Theoretical pH-titration curve of an ionizable residue. The chemical shift evolution as a function of pH is collected for a particular nucleus (^{1}H here). At low pHs (left), the acidic form of the amino acid is the major form, while the basic form is preponderant at high pHs. The inflection point of the curve corresponds to equal acidic and basic populations and pH = pKa.

To determine the pKa of ionizable residues in a protein, it is important to follow the chemical shifts of several nuclei through 2D and/or 3D-NMR experiments. ^1H_N and ^{15}N amide chemical shifts can be monitored in 2D ^{1}H-^{15}N-HSQCs. The ^{15}N shifts of ionizable residues are dominated by the protonation of the residue's own side-chain (Croke *et al.*, 2011). Aliphatic ^{1}H and ^{13}C chemical shifts can be followed in ^{1}H-^{13}C-HSQCs, while $^{2-3}J_{H-N}$ ^{1}H-^{15}N-HMQCs and aromatic ^{1}H-^{13}C-HSQCs will allow to monitor the side-chains of His and Tyr directly involved in the protonation state changes.

To study Asp and Glu residues in uniformly (^{15}N)-^{13}C-labeled proteins, HCACO optimized for the detection of aspartyl Cγ and glutamyl $^{13}C_\delta$, and HCCH-TOCSY are performed to direclty follow the side-chains carbonyl resonances of these two residues (Everill *et al.*, 2012). For Lys and Arg residues, Andre *et al.* described a 3D HNC experiment that involves $^2J_{H\epsilon-N\zeta}$ and $^2J_{H\delta-N\epsilon}$ coherence transfer in

the Lys and Arg side-chains, respectively (André *et al.*, 2007). Since non-exchangeable protons are used for excitation and detection, these experiments are applicable over the full range of pH and temperature. These two nitrogen nuclei are suitable reporters for Arg and Lys charge states, as $N\epsilon$ for Arg is involved in the charge delocalization over the $N\epsilon - C\zeta - (NH_2)_2$ moiety, and as $N\zeta$ in the case of Lys is directly bound to the titrating proton.

For cysteine, the sulfur atom is usually not observed with standard NMR probes. It is thus not possible to detect the nucleus directly bound to the titrating proton. The pKa determination relies on the lone N^H, Cα-Hα and Cβ-Hβ chemical shift through a ^{1}H-^{15}N-HSQC and a ^{1}H-^{13}C-HSQC centered on aliphatics. For His and Cys, the use of 3D-NMR is therefore not required, as involved nuclei are often well-resolved and can be easily followed along the titration.

4.2.3 Titration curve fitting and pKa determination

When it comes to the pKa determination of an amino acid by NMR, the pH-dependence of the chemical shift for each of the sensitive nuclei of the targeted amino acid must be individually analyzed through a three-parameter fit with a following equation:

$$\delta_{obs} = \frac{\delta_A + \delta_B 10^{(pH-pKa)}}{1 + 10^{(pH-pKa)}} \tag{4.14}$$

where δ_A (the chemical shift of the acidic form), δ_B (the chemical shift of the basic form) and pKa are the unknown to determine. The pKa value can only be assessed to be correct if the chemical shift jump occurs within 4 pH units and if obtained pKa value is the same for all independently-fitted nuclei. This is generally the case when the concerned ionizable residue is isolated from others, as chemical shifts are influenced by their environment within 5-10 Å. This situation is often rare in particular for histidine residues, as shown in figure 4.7 in the case of Ldt_{Bs}. The experimental data points obtained for the His 122 Hδ2 along the titration show a chemical shift jump on a pH-window that is larger than 4 units (figure 4.7a). Furthermore the shape of the titration curve for the Cδ2 is not consistent with the theoretical curve presented in figure 4.6. This could indicate the presence of an ionizable residue in the vicinity of His 122, with a pKa that is within 4 units to the actual pKa value of His122. To test this hypothesis a Hill coefficient can be introduced in the fitting equation to take into account this cooperativity between two pKas. The fitting equation becomes:

$$\delta_{obs} = \frac{\delta_A + \delta_B 10^{n(pH-pKa)}}{1 + 10^{n(pH-pKa)}} \tag{4.15}$$

where n is the Hill coefficient. When close to 1, the titration site does not interact with neighboring ionizable groups, whereas a n below unity is the signature of the presence of other ionizable residues in the vicinity of the residue of interest.

In cases where n is lower than 1 and where a structure is available, fits with two or three pKas can be undertaken with the following equations (see also figure 4.7, parts b and e):

$$\delta_{obs} = \frac{\delta_A 10^{-2pH} + \delta_B 10^{-(pH+pKa)} + \delta_C 10^{-(pKa+pKa_1)}}{10^{-2pH} + 10^{-(pH+pKa)} + 10^{-(pKa+pKa_1)}} \tag{4.16}$$

$$\delta_{obs} = \frac{\delta_A 10^{-3pH} + \delta_B 10^{-(2pH+pKa_1)} + \delta_C 10^{-(pH+pKa+pKa_1)} + \delta_D 10^{-(pKa+pKa_1+pKa_2)}}{10^{-3pH} + 10^{-(2pH+pKa_1)} + 10^{-(pH+pKa+pKa_1)} + 10^{-(pKa+pKa_1+pKa_2)}} \quad (4.17)$$

where pKa is still the pKa value of the residue of interest, whereas pKa_1 and pKa_2 can represent either close ionizable residues pKa or the pKa of the buffer (like phosphate that was used in the case of Ldt_{Bs} studies). In this case δ_A represents the chemical shift of the acidic form of the residue of interest when the neighboring ionizable groups are also in their acidic form, δ_B, δ_C and δ_D represent the chemical shifts of the successive acido-basic forms of the residue of interest in contact with the different ionization states of the considered neighboring residues.

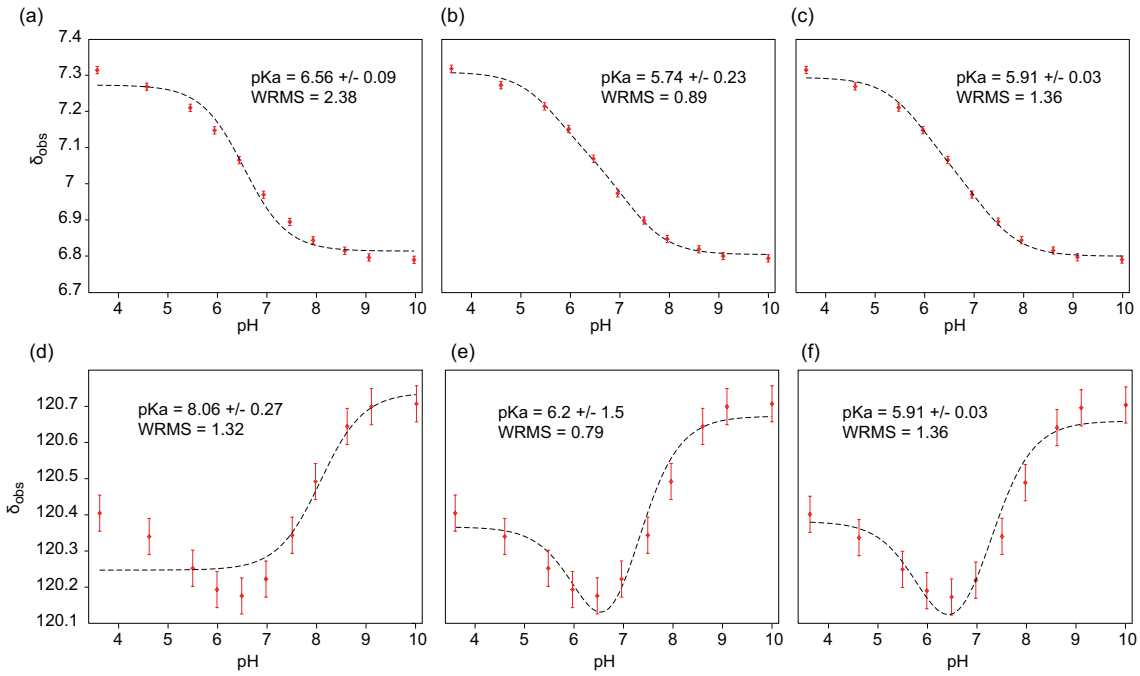


Figure 4.7: Evaluation of different fitting procedures of the titration curves of residue His122 Hδ2 (a-c) and Cδ2 (d-f) nuclei in Ldt_{Bs}. Chemical shifts are fitted individually with one pKa according to equation 4.14 (a, d), with two pKas according equation 4.16 (b, e) and with all other nuclei of the residue with two pKas still according equation 4.16 (c, f). In the last fitting procedure, one of the pKas is supposed to arrive from the phosphate buffer due to the solvent exposure of His122 and the absence for neighboring titrating residues, the second pKa is fitted freely within the procedure. The pKa values obtained from the fitting procedures are given in each case, together with the WRMS (Weighted Root Mean Square Deviation) that comes from the Monte-Carlo protocol within the modelXY procedure of NMRPipe (the routine used to perform the fit). The WRMS value reflects the quality of the performed fit (the lower this value, the better the quality of the fit).

As the number of fitted-parameter increase, it becomes impossible to perform well with a limited number of data points. Advantage must then be taken from the diversity of probes on the same residue. For this reason, H_N, Hα, Hβs, Hδ2, Hϵ1, Cα, Cβ, Cδ2, Cϵ1, N, Nδ1, Nϵ2 experimental chemical shifts for all measured values need to be included in a fitting procedure with single pKa values. The benefit that can be obtained from this approach is illustrated in figure 4.7c where a two-pKa fit has been performed for His122 of Ldt_{Bs}. Not only did this approach improved the fitting quality for the nucleus Cδ2, but the obtained pKa values are uniform all over the nuclei. Surprisingly, this approach was never performed,

to our knowledge, with such a high number of NMR probes in the literature, thus increasing drastically the confidence in the pKa determination. Rather pKa values are determined on a small number of nuclei, aromatic carbons or aromatic nitrogens in general, through individual fits.

The main result of this study is that the β position is the less sensitive to the proximity of ionizable groups. Indeed, the pKa obtained for His residues in Ldt_{Bs} by the complete approach, gave similar values to the simple 2 pKa fits on individual Hβs and Cβ nuclei. Among the aromatic nuclei the Cδ2 is by far the most sensitive to the ionization of proximal groups, as shown by the chemical shift dependence of this nucleus with pH in figure 4.7d-f.

The present study also reveals the importance of the choice of the buffer for pH titrations, as buffer molecules might interact with solvent exposed ionizable residues, as apparently shown here for Ldt_{Bs} His122 and phosphate ions from the buffer.

4.2.4 The particular case of histidines

4.2.4.1 Exploring histidine protonation states

The histidine imidazole ring

Histidine is an essential residue in proteins, as $\sim 50\%$ of enzymes contain histidines in their active sites, where they are involved in catalysis, pH regulation (Hass *et al.*, 2008), phosphorylation (Dutta *et al.*, 1999), and metal binding, as in zinc finger motifs (Jensen *et al.*, 2007) for example. Among the twenty residues found in proteins, histidines are the most involved in catalytic reactions (figure 4.5). It differs from the previously described residues as its imidazole ring side-chain can exist in three different protonation states, including an acidic fully protonated, charged imidazolium and a pair of basic neutral tautomers (figure 4.8a), the Nδ1-H and the Nϵ2-H tautomers.

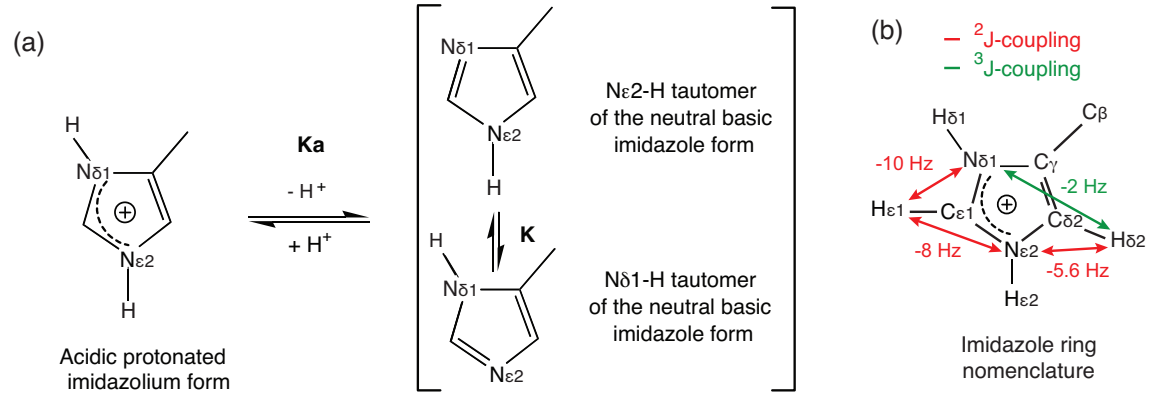


Figure 4.8: Description of the protonation states of histidine imidazole rings. **(a)** Representation of the acido-basic and tautomeric equilibria involving the histidine imidazole ring. Instead of having one acid and one basic form, histidine shows one acid (fully protonated) and two neutral basic forms (the two tautomers). **(b)** $^{2}J_{H-N}$-couplings (red) and $^{2}J_{H-N}$ couplings (green) detected in ^{1}H-^{15}N-HMQC experiments usually used to determine histidine protonation states. Approximate values of these couplings are indicated and were taken from (Blomberg *et al.*, 1977).

In the free amino acid, in peptides or in water-accessible areas of a protein, the Nϵ2-H tautomer is usually the most stable, preponderant basic form. On the contrary, when the histidine is buried and involved in hydrogen bond networks, the basic imidazole ring can exist primarily as the Nδ1-H-tautomer.

This equilibrium between the two tautomeric basic forms can have a role in the regulation of the activity of the protein. In the neutral form, the non-protonated nitrogen can serve as an effective ligand for metal binding, or can play a crucial role in proton-transfer processes (Hass *et al.*, 2008). pH and protein environment thus cooperatively condition the stability of a particular protonation state.

The NMR chemical shifts (^{15}N-chemical shifts in particular) are very sensitive to the protonation and tautomeric states of histidines. NMR experiments (in both liquid- (Pelton *et al.*, 1993) and solid-state (Li and Hong, 2011)) have thus been widely developed to enable the identification of these histidine protonation states.

Nitrogen chemical shifts in the imidazole ring

^{15}N-chemical shift of nitrogen nuclei within the imidazole ring can be indirectly investigated through a ^{1}H-^{15}N-Heteronuclear Multiple-Quantum Coherence (HMQC) experiment optimized for the detection of $^{2}J_{NH}$ and $^{3}J_{NH}$ couplings (figure 4.8b) (Pelton *et al.*, 1993). The Nδ1-Hϵ1, Nδ1-Hδ2, Nϵ2-Hϵ1 and Nϵ2-Hδ2 correlations are detected at characteristic positions and relative intensities for each individual form, as depicted in figure 4.9. To assign the four correlation peaks arising from each histidine, both chemical shifts and peaks intensity will be helpful. First, the proton dimension allows the assignment of Hδ2 and Hϵ1 according to their average chemical shift value, at 7.01 ppm and 7.95 ppm, respectively (values taken from the BMRB). Second, intensities are directly correlated to the scalar coupling constants, according to:

$$\text{peak intensity} \sim \sin^2(\pi J \tau) \tag{4.18}$$

where τ is the mixing period during which ^{1}H and ^{15}N signals become antiphase, set to 22 ms. As the $^{2}J_{NH}$ scalar coupling constants are stronger than the $^{3}J_{NH}$ coupling constant (figure 4.8), the Nδ1-Hδ2 correlation is expected to show lower intensity than the correlations arising from N and H nuclei separated by two covalent bonds.

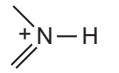
Structure			
Nitrogen type	α+	α	β
Chemical shift (ppm)	176.5	167.5	249.5

Table 4.2: ^{15}N chemical shifts expected for nitrogens depending on their protonation and hybridization state. Chemical shift values for these nitrogen types have been assembled from various model compounds in aqueous solutions (Pelton *et al.*, 1993; Day *et al.*, 2003).

Taking into account the typical ^{15}N-chemical shift values (Table 4.2) and $^{2}J_{NH}$ / $^{3}J_{NH}$ scalar coupling parameters, different characteristic ^{1}H-^{15}N-HMQC patterns are expected for each of the protonation and tautomeric states of the histidine imidazole ring. These patterns are presented in figure 4.9. The close proximity of Nδ1 and Nϵ2 chemical shift unambiguously identifies the acidic protonated form. The low intensity of the Nδ1-Hδ2 correlation establishes the chemical shift of the Nδ1 nitrogen in the neutral basic form and its value discriminates the two possible tautomers. As indicated in Table 4.2, the chemical shift of Nδ1 and Nϵ2 will be inverted depending if there are in a type-α (167.5 ppm) or a type-β configuration (249.5 ppm).

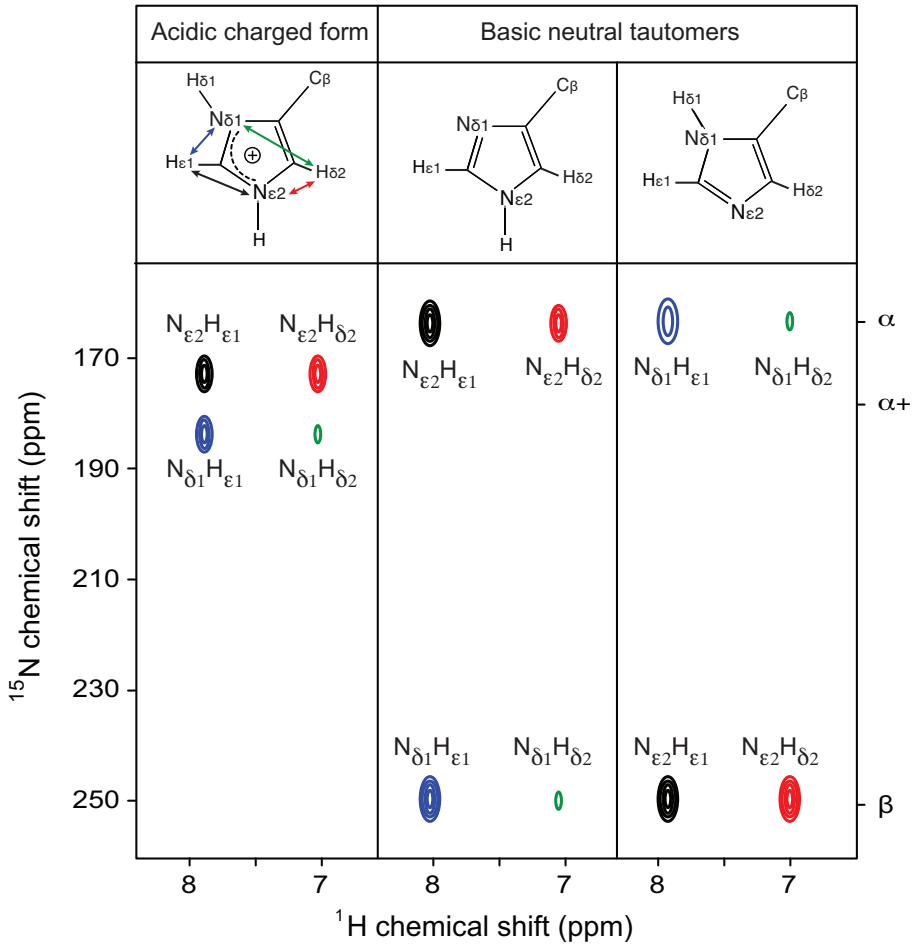


Figure 4.9: Expected ^{1}H-^{15}N-HMQC NMR spectra for the three individual protonation states of the histidine imidazole ring. For the acidic charged form, Nδ1 and Nϵ2 chemical shifts are expected to be very close, whereas they will be significantly spread out for the two neutral forms. The chemical shift of the nitrogen involved in the double-bond (type-β) is expected at ∼ 250 ppm and the one involved in the single bond (type-α) at ∼ 170 ppm. Figure adapted from (Pelton *et al.*, 1993).

The situation in proteins may be more complex. First the observed ^{15}N chemical shift of each protonation and tautomeric state of the imidazole ring can be influenced by the histidine local environment in the protein, by electrostatic charges and hydrogen-bond networks. This effect has been estimated to 10 ppm at the maximum. At high pHs, the basic histidine can exist as a mixture of the two tautomers in fast equilibrium, leading to an intermediate chemical shift. This situation may lead to the estimation of the corresponding equilibrium constant K (figure 4.8a). It was encountered in the case of Ldt_{fm} His421, an histidine located at the active site.

In cases where exchangeable protons Hδ1 or Hϵ2 in imidazole rings are protected from solvent accessibility or involved in hydrogen-bond networks, they can potentially be identified and assigned in ^{1}H-^{15}N-HSQC ($^{1}J_{NH}$ ∼ 100 Hz) or ^{1}H-^{15}N-HMQC spectra (with the tau value set to 4.5 ms). They usually have a typical ^{1}H chemical shifts above 11.5-12 ppm (Slupsky *et al.*, 1998). Such protons have been detected in Ldt_{Bs} and Ldt_{fm} in histidine involved in the active site, *i. e.* His^{126} and His^{421}, respectively. Their identity has been established through the covalently bound ^{15}N-nucleus.

Carbon chemical shifts in the imidazole ring

As well as nitrogen nuclei, carbons can be used to discriminate between His tautomers. ^{13}C NMR chemical shifts can be detected through a constant-time ^{1}H-^{13}C-HSQC centered at 127 ppm in the indirect ^{13}C-dimension, with $^{1}J_{C-H} \sim 160$ Hz and $^{1}J_{C-C} \sim 60$ Hz. One of the first investigations of His tautomeric forms by ^{13}C NMR was performed in 1973 on a histidine methylated at positions Nδ1 or Nϵ2, and revealed a chemical shift difference for Cδ2 of 9.2 ppm between the two tautomers, but no difference for Cϵ1 (Reynolds *et al.*, 1973).

In proteins, a recent comparison of numerous ^{13}C histidine chemical shifts determined by both liquid- and solid-state NMR data deposited in the BMRB (Vila *et al.*, 2011) confirmed that Cϵ1 are not sensitive to the histidine tautomer, whereas Cδ2 and Cγ are strongly affected by the changes in the electronic distribution within the imidazole ring. The authors showed two contributions in the chemical shifts distribution for Cδ2, a major one centered at 119 ppm and a minor one at 127 ppm (on 2,267 chemical shifts in total, figure 4.10a). These values are close to the 9.2 ppm difference shown for the methylated histidine. Cγ also seemed to exhibit two contributions, but the low number of available data (84 chemical shifts) is not sufficient to comment on this nucleus. Cϵ1 only showed a single contribution, centered at 140 ppm (figure 4.10b).

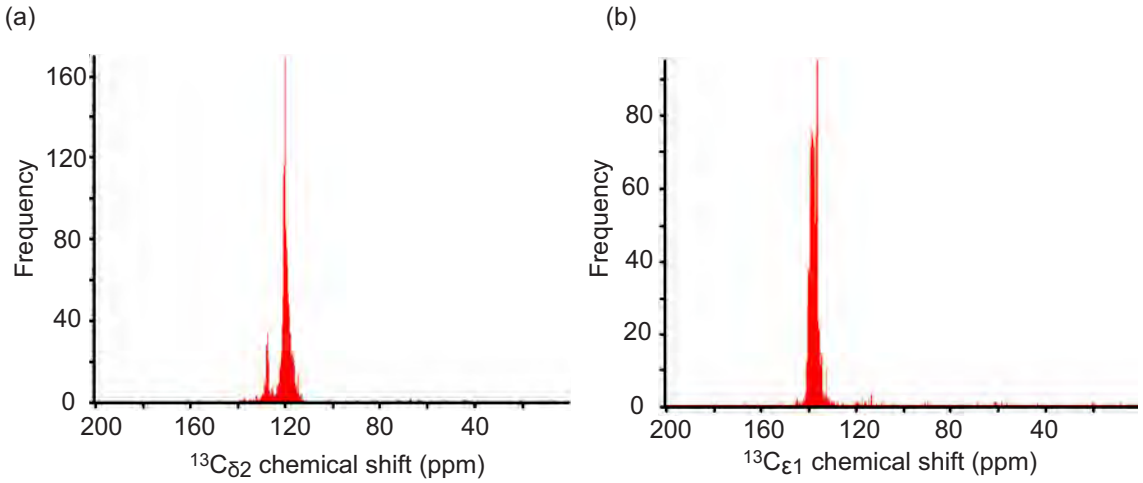


Figure 4.10: Histogram of the histidine ^{13}C observed chemical shift distribution for a) Cδ2 and b) Cϵ1. Data were obtained from the BMRB and contained 2,267 values for Cδ2 and 1,754 for Cϵ1. Cδ2 shows two major contributions, whereas Cϵ1 shows a single mono-disperse distribution. Adapted from supporting information of Vila *et al.* (Vila *et al.*, 2011).

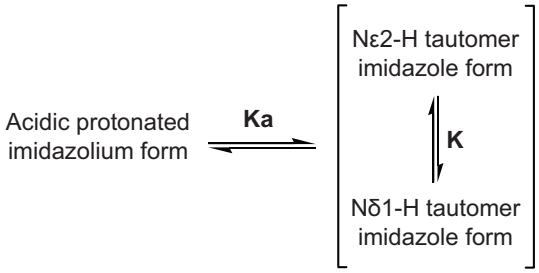
Vila *et al.* suggested that the minor population of histidine Cδ2 centered at 127 ppm contains mainly contributions from the Nδ1-H tautomer, whereas the major fraction centered at 119 ppm contains contributions from both the acidic charged form and the Nϵ2-H tautomer (Vila *et al.*, 2011). On the basis of this empirical analysis, the authors developed a new computational approach enabling to determine the tautomeric fractions of the imidazole ring of histidines in proteins as a function of pH. This method requires the observed Cγ and Cδ2 chemical shifts and either the structure of the protein, or the fraction of the protonated form.

This BMRB database investigation on proteins provides interesting view, as it reveals that the Nδ1-H tautomer can be identified thanks to the lone histidine Cδ2 chemical shift, as it was previously indicated

in a ^{13}C-NMR study (Sudmeier *et al.*, 2003). Contrary to nitrogen chemical shifts, which enable to easily differentiate between the acidic from the basic forms of histidines, carbon chemical shifts can provide a complementary information to differentiate the Nδ1-H tautomer and the two other forms of the imidazole ring. In the case of LDts, we observed this typical Cδ2 chemical shift revealing the presence of the Nδ1-H tautomer in free and acylated Ldt$_{\mathrm{Bs}}$, as well as in Ldt$_{\mathrm{fm}}$ acylenzyme. For Ldt$_{\mathrm{fm}}$ apoenzyme, the occurence of the Nδ1-H tautomer is pH-dependent.

Nδ1, Nϵ2 and Cδ2 are thus key nucleic probes to investigate the protonation and tautomeric states of histidine residues within proteins. With the determination of the combined pKa and K thermodynamic constants, a precise view of the distribution of the three forms of the imidazole ring at a given pH can be obtained, thus leading to a better understanding of the implication of histidines in catalytic mechanisms.

4.2.4.2 Distribution of the three histidine acido-basic and tautomeric forms at a given pH



Under the assumption that the acido-basic and the tautomeric equilibria are independent and that the three forms are experimenting a fast exchange regime, the pKa and the constant K (reminded above) can be determined. pKas are extracted from the titration curves fitted simultaneously on all sensitive nuclei (optimally H_N, Hα, Hβs, Hδ2, Hϵ1, Cα, Cβ, Cδ2, Cϵ1, N, Nδ1 and Nϵ2). The constant K is determined from the extrapolated chemical shift values at basic pHs of the Nδ1, Nϵ2 and Cδ2 chemical shifts.

For these three sensitive imidazole ring nuclei, the fraction of the Nϵ2-H tautomer is calculated independently. The following equations were deduced from (Reynolds *et al.*, 1973) for the Cδ2, and from (Bachovchin, 1986) for the Nδ1 and Nϵ2, and were used in the case of LDts.

Calculation of the Nϵ2-H tautomer fraction ($f(\mathrm{N}\epsilon2_{\mathrm{C}\delta2})$) from the C$\delta$2 chemical shift:

$$f(\mathrm{N}\epsilon2_{\mathrm{C}\delta2}) = \frac{(\delta_{obs\ C\delta2-A} - \delta_{obs\ C\delta2-B}) - \Delta_{C\delta2,N1-His}}{\Delta_{C\delta2,N3-His} - \Delta_{C\delta2,N1-His}} \tag{4.19}$$

where $\delta_{obs\ C\delta2-A}$ and $\delta_{obs\ C\delta2-B}$ are the experimental Cδ2 chemical shifts at acidic and basic pHs deduced from the fit of the titration curve, $\Delta_{C\delta2-N1-His}$ (value -7.1) is the Cδ2 chemical shift difference measured between the acidic and basic forms of 1-mehylhistidine (also called Nϵ2-methylhistidine) and $\Delta_{C\delta2-N1-His}$ (value 2.1) is the Cδ2 chemical shift difference measured between the acidic and basic forms of 3-mehylhistidine (also called Nδ1-methylhistidine).

Calculation of the Nϵ2-H tautomer fraction from the Nδ1 ($f(\mathrm{N}\epsilon2_{\mathrm{N}\delta1})$) and from the N$\epsilon$2 ($f(\mathrm{N}\epsilon2_{\mathrm{N}\epsilon2})$) chemical shift:

$$f(\mathrm{N}\epsilon 2_{\mathrm{N}\delta 1}) = \frac{\delta_{obs\ N\delta 1} - \delta_{\alpha}}{\delta_{\beta} - \delta_{\alpha}} \tag{4.20}$$

$$f(\mathrm{N}\epsilon 2_{\mathrm{N}\epsilon 2}) = \frac{\delta_{obs\ N\epsilon 2} - \delta_{\beta}}{\delta_{\alpha} - \delta_{\beta}} \tag{4.21}$$

where $\delta_{obs\ N\delta 1}$ and $\delta_{obs\ N\epsilon 2}$ are the experimental Nδ1 and Nϵ2 chemical shifts at basic pH deduced from the fit of the titration curve, and δ_{α} and δ_{β} are the chemical shifts reported in Table 4.2 (page 118) for nitrogens of type α (δ_{α} = 167.5 ppm) and β (δ_{β} = 249.5 ppm), respectively.

The final fraction of Nϵ2-H tautomer $f_{N\epsilon 2}$ is then determined as the average of these three values (or one or two of them depending on the available experimental data). The deviations between these values are used to estimate errors and to evidence discrepancies, that could come from interdependence between the acid-base and tautomerism equilibria.

The equilibium constant K is finally calculated as follows:

$$K = \frac{1 - f_{N\epsilon 2}}{f_{N\epsilon 2}} \tag{4.22}$$

The distribution of each of the three histidine forms at a given pH is then calculated from the pKa and K values determined previously:

$$f_{\,\mathrm{acid-form}} = \frac{1}{1 + 10^{(\mathrm{pH-pKa})}} \tag{4.23}$$

$$f_{\,\mathrm{N}\epsilon 2-\mathrm{H-tautomer}} = \frac{1 - f_{\,\mathrm{acid-form}}}{1 + \mathrm{K}} \tag{4.24}$$

$$f_{\,\mathrm{N}\delta 1-\mathrm{H-tautomer}} = \frac{1 - f_{\,\mathrm{acid-form}}}{1 + \dfrac{1}{\mathrm{K}}} \tag{4.25}$$

Description of the protonation states of amino acids and of their potential distribution between different forms at different stages of a reaction pathway brings essential information to decipher the interaction mechanism between a protein and its natural ligand or competitive inhibitors. It is a nice addition to the thermodynamic parameters, such as enthalpy, entropy or dissociation constants that can also be determined by NMR, especially in the case of low affinity complexes. These data nevertheless offer a static picture of the reaction pathway. NMR has clearly more to offer with a dynamic view as described in the next chapter.

Chapter 5

Dynamical approach by NMR

When used as a structural biophysical technique, NMR can provide key information on the ground-state of three dimensional structures. When it comes to understand the relationship between structure and function, additional information is required on excited states (Kay, 1998). A number of studies have proven that there is an intimate relation between protein dynamics and molecular function, including ligand recognition and binding when "closed" regions are required to open for the entrance of the molecule, catalysis where active site residues can encounter conformational rearrangements and allostery (Mittermaier and Kay, 2006). NMR can play a major role in this context, in part due to the sampling of a broad range of dynamics time-scales (figure 5.1).

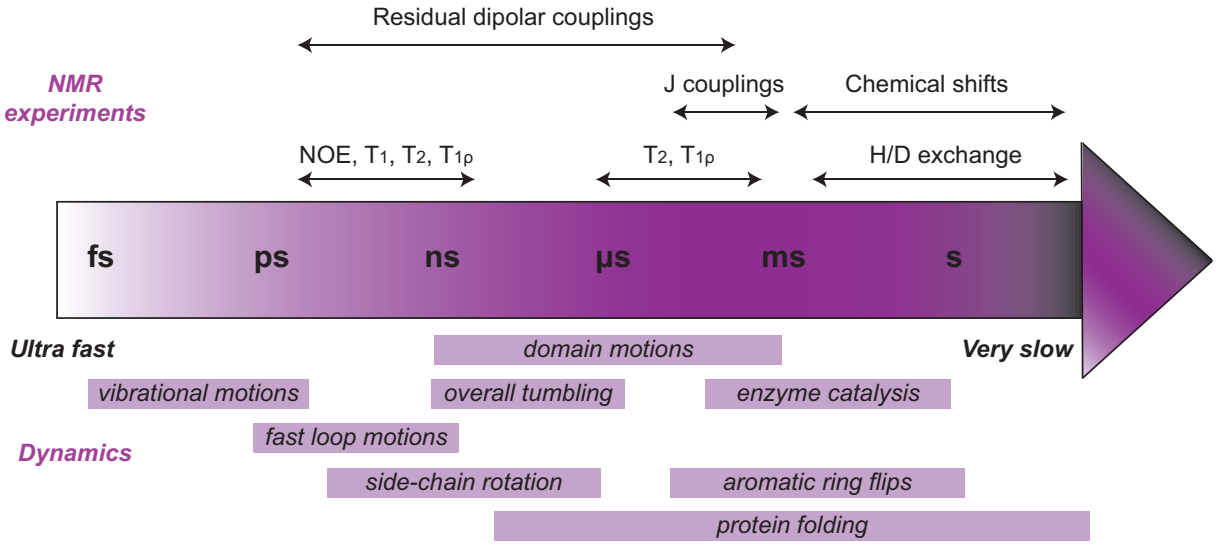


Figure 5.1: Overview of protein dynamics on different time-scales. On top of the arrow, NMR parameters are indicated with their corresponding time-scale. Below the arrow are found the biological processes that are believed to involve the sampled dynamics events.

A large number of NMR tools have been developed in order to be able to characterize these dynamic processes in proteins and to identify the time-scale in which the dynamic processes occur. NMR spectroscopy is uniquely suited to these studies as it provides site-specific information. The dynamic processes cover a large time-scale regime, ranging from ultrafast movement (fentosecond for the molecular internal vibrations), to very slow motions (chemical shifts or Hydrogen/Deuterium H/D exchange, typically slower than second) (figure 5.1). Dynamics is therefore one of the fundamental aspects of NMR, and is complementary to thermodynamical and structural approaches described in the preceding chapters.

During the inhibition study of Ldt_{Bs} by β-lactams, it clearly appeared that the enzyme was affected by flexibility upon ligand binding, as many resonances were disappearing from the spectrum of the acylenzyme. Signal disappearance can be the signature of relaxation phenomenon. We thus decided to investigate motions in the ns-ms time-scales in this class of enzymes, with focus on amide ^{15}N nuclei.

The exploration of the ^{15}N-relaxation offers information on the time-scale of global and local molecular reorientation of the ^{1}H-^{15}Nvector, in other words information on backbone dynamics.

The present chapter briefly details the main principles of the study of protein dynamics by NMR relaxation. The relaxation parameters give access to two different motions: fast motions at the pico- to nanosecond time-scale through T_1, T_2 and NOE parameters, and slow motions at the micro- to millisecond time-scale through T_2 and $T_{1\rho}$ parameters. Relaxation phenomena are described, with emphasis on practical methods that have been used to extract LDts relaxation rates and time-scale of motions. For more details on the theoretical analysis of NMR relaxation, the reader can refer to several books and reviews (Fischer *et al.*, 1998; Palmer, 2001; Kleckner and Foster, 2011).

5.1 NMR relaxation: origin and measurement of characteristic parameters

In the presence of a magnetic field B_0, the nuclear spins are distributed on different energy levels, corresponding to the thermodynamical equilibrium. When a perturbation is applied to a system, by for instance an excitation induced by a magnetic field B_1, spins are driven towards higher energy levels thus perturbing the initial population distribution. The system will naturally return to equilibrium by a process called relaxation. Nuclear relaxation is induced by different mechanisms, among which the dipolar interactions, the Chemical Shift Anisotropy (CSA) and the quadrupolar interactions. These phenomena induce local magnetic fields at the protein nuclei, that are fluctuating due to the global and internal movements of the protein. Dipolar interactions occur between the nuclear magnetic moments and is the most important contribution to relaxation in biological macromolecules. CSA results from the interaction of the nuclei with the local magnetic field and reflects the sampling of different orientations by the molecule. Quadrupolar interactions only occur for nuclear spins $I > 1$ and are thus null when ^{1}H, ^{13}C and ^{15}N are concerned.

NMR relaxation involves two distinct processes, the spin-lattice relaxation T_1, also called longitudinal relaxation, and the spin-spin relaxation T_2, also called transverse relaxation. T_2 determines the peaks line-width in the NMR spectra and T_1 fixes the fastness at which an experiment can be repeated. These parameters depend on the correlation time τ_c of the molecule and on local motions. τ_c represents the average time for the molecule to rotate by one radian, and it varies with molecular size, sample viscosity and temperature. The dependence to τ_c is coded in the spectral density function $J(\omega)$, that represents the probability for a ^{15}N-^{1}H vector to sample a molecular motion at a given angular frequency ω (in rad.s^{-1}).

5.1.1 Longitudinal relaxation

5.1.1.1 Origin of longitudinal relaxation

Longitudinal relaxation is the mechanism by which an excited magnetization vector returns to equilibrium along the external magnetic field B_0, conventionally oriented on the z-axis (figure 5.2a). This process follows a pseudo-first order kinetics characterized by a constant time T_1 (or a rate constant R_1 = $1/T_1$):

$$\frac{dM_z}{dt} = \frac{M_{z,eq} - M_z}{T_1} \tag{5.1}$$

An example of this relaxation process is illustrated in figure 5.2a, where the equilibrium magnetization M_0 is initially inverted and follows a mono-exponential time-dependence to recover its value at equilibrium (figure 5.2b):

$$M_z(t) = M_{z,eq}\left(1 - 2e^{-\frac{t}{T_1}}\right) \tag{5.2}$$

where $M_z(t)$ is the magnetization along the z axis at time t.

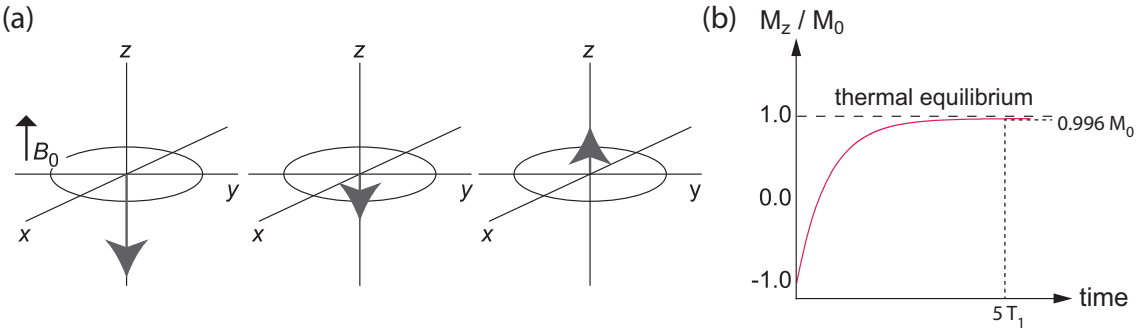


Figure 5.2: Time-evolution of the magnetization with longitudinal relaxation. **a)** Vector diagram of the evolution of the M_z magnetization at different times after inversion of the equilibrium magnetization. **b)** Time-evolution of M_z/M_0. The z-magnetization changes its sign for $T_1 = \ln 2$ and is recovered at 99.6 % after 5 T1. The latter result conditions the fastness at which an NMR experiment can be repeated.

5.1.1.2 Impact of molecular motion

^{1}H longitudinal relaxation is dependent on the molecular motion, through the spectral density function. The T_1-dependence to the molecular correlation time τ_c is illustrated in figure 5.3. When τ_c increases, T_1 becomes sensitive to the magnetic field and there is an advantage to work at lower field with that respect for solids in particular.

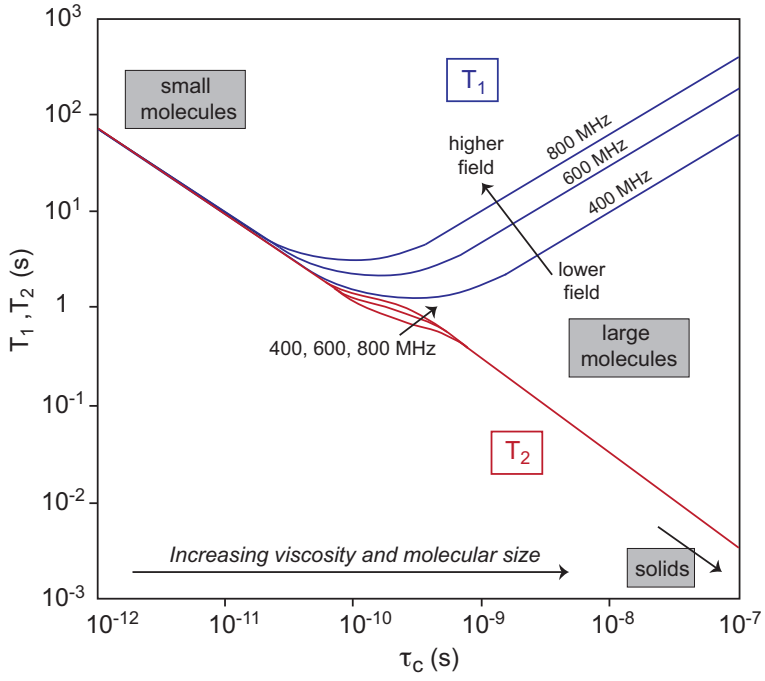


Figure 5.3: Evolution of the relaxation time-constants T_1 and T_2 with the correlation time τ_c of the molecule. T_2 decreases continuously as the size of the molecule increases, thus causing line-broadening and signal-detection hampering for large molecules (linewidth at half-height is $1/(\pi T_2)$. T_1 is proportional to B_0^2 for proteins.

5.1.1.3 ^{15}N-longitudinal relaxation: measurement and dependence to motion

^{15}N-longitudinal relaxation rates are usually measured using a modified ^{15}N-^{1}H-HSQC which allows to create an inverted Nz magnetization and to measure its evolution during a variable delay t before to

edit the nitrogen frequency (Farrow *et al.*, 1994). During the relaxation delay, ^{1}H 180° pulses are applied to suppress the effects of cross-correlation. A phase cycle is applied to get rid of the dependence with the initial magnetization and the intensity of each ^{1}H-^{15}N correlation follows a mono-exponential decay with a time constant T_1:

$$I(t) = I_0 \, e^{(-R_1 t)} \tag{5.3}$$

where $I(t)$ represents the amide cross-peak intensity for the delay t and I_0 the cross-peak intensity at $t = 0$. Figure 5.4 shows an example of the determination of ^{15}N-R_1 in the case of the amide resonance of residue 62 in Ldt_{Bs}.

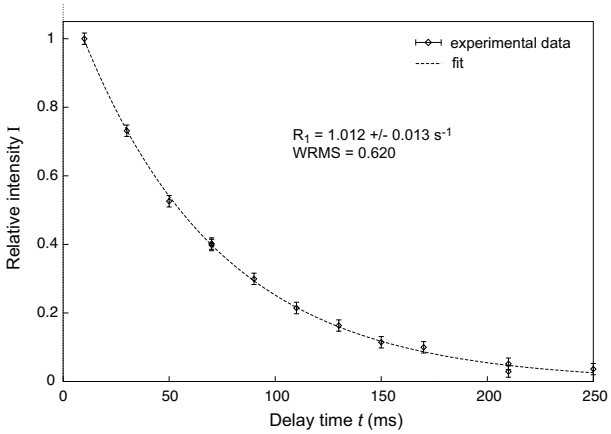


Figure 5.4: Determination of the ^{15}N-relaxation rate R_1 for the amide resonance of residue 62 in Ldt_{Bs}. ^{15}N-^{1}H-HSQCs are collected with variable t delays to allow Nz return to equilibrium after inversion. Intensity of the cross-peak is measured as a function of t and reported on this graph. The curve is then fitted with equation 5.3 to extract R1. The error on R_1 is evaluated through a Monte-Carlo analysis and the quality of the fit is underlined by the weighted root-mean squared deviation (WRMSD).

The ^{15}N-R_1 relaxation rate is dependent on molecular motions, as emphasized through the following equation that involves the spectral density function (Kneller *et al.*, 2002):

$$R_1 = \frac{d^2}{10}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{2}{5}c^2 J(\omega_N) \tag{5.4}$$

The contribution of dipolar interaction is shown in red while the contribution of the CSA is shown in green in the equation. $J(\omega)$ is the spectral density function, which can be described by a Lorentzian in the case of a spherical molecule (isotropic conditions):

$$J(\omega) = \frac{2}{5}\left[\frac{\tau_c}{1 + (\omega\tau_c)^2}\right] \tag{5.5}$$

where ω is the Larmor angular frequency of the nucleus of interest (^{15}N or ^{13}C), and with:

$$\tau_c = \frac{1}{6D_{iso}} \qquad with \;\; D_{iso} = \frac{kT}{8\pi\eta r^3} \tag{5.6}$$

where D_{iso} is the isotropic rotational diffusion constant, η is the viscosity, k the Boltzmann constant, T the temperature and r the radius of the molecule. d and c from equation 5.4 are the dipolar interaction constant and the CSA constant, respectively, defined by:

$$d = \frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2 r_{NH}^3} \tag{5.7}$$

and

$$c = \frac{\omega_N \Delta\sigma_N}{\sqrt{3}} \tag{5.8}$$

where ω_H and ω_N represent the Larmor angular frequencies for the ^{1}H and ^{15}N nuclei, respectively; γ_H and γ_N are the respective gyromagnetic ratio; μ_0 is the vacuum permeability, h is the Planck's constant, r_{NH} is the interatomic ^{1}H-^{15}N distance (1.02 Å); and $\Delta\sigma_N = \sigma_{/\!/} - \sigma_\perp = -172 \times 10^{-6}$ with $\sigma_{/\!/}$ and $\sigma_\perp$ the parallel and perpendicular components of the ^{15}N chemical shift tensor, respectively.

5.1.2 Transverse relaxation

5.1.2.1 Origin of transverse relaxation

Transverse relaxation T_2 is the second relaxation process and is the mechanism by which the excited magnetization vector decays in the x-y plane, perpendicular to the magnetic field B_0 (figure 5.5a). This process follows a pseudo-first order kinetics characterized by a constant time T_2 (or a rate constant R_2 = 1 / T_2):

$$M_{x,y}(t) = M_{x,y}(0)\, e^{-\frac{t}{T_2}} \qquad with\ \ M_{x,y} = \sqrt{M_x^2(t) + M_y^2(t)} \tag{5.9}$$

where M_x and M_y are the x and y components of the transverse magnetization. Transverse relaxation is equal (small molecule, extreme narrowing) or faster ($T_2 < T_1$) than the longitudinal relaxation (large molecules).

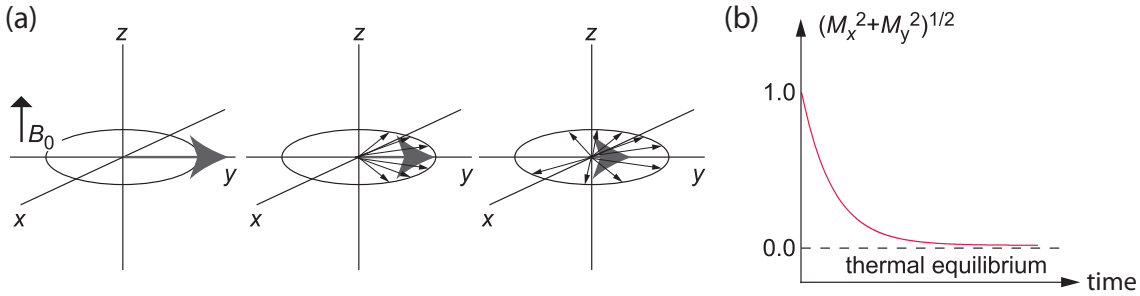


Figure 5.5: Time-evolution of the magnetization with transversal relaxation. a) Vector diagram of the evolution of the magnetization in the x-y plane after excitation of M_z through a 90° pulse. The individual spins (small arrows) are gradually dephasing with respect to one another (loss of coherence), leading to a decrease of the macroscopic in-plane magnetization (large arrow). **b)** Time-evolution of the in-plane magnetization. The in-plane magnetization follows a mono-exponential decay. Only 0.7 % of the initial magnetization is left after 5 T_2. T_2 thus fixes the decay of the free induction signal after excitation (FID).

5.1.2.2 Impact of molecular motion

^{1}H transversal relaxation is dependent on the molecular motion, through the spectral density function. The T_2-dependence to the molecular correlation time τ_c is illustrated in figure 5.4. As τ_c increases, T_2 diminishes, yielding to very broad linewidth in the NMR spectra that severely hamper the detection of

the signal in proteins over 50 kDa. The main origin of this phenomenon is the ^{1}H-^{1}H dipolar interaction that can be reduced by the deuteration of the protein (γ_H/γ_D = 6).

5.1.2.3 ^{15}N-transverse relaxation: measurement and dependence to motion

^{15}N-transversal relaxation rates are usually measured using a modified ^{15}N-^{1}H-HSQC which allows to create a Nx or Ny magnetization and to measure its evolution during a variable delay t before to edit the nitrogen frequency (Farrow *et al.*, 1994). In a similar fashion to what was described in section 5.1.1.3, the intensity of each ^{1}H-^{15}Ncorrelation follows a mono-exponential decay with a time constant T_2:

$$I(t) = I_0\, e^{(-\frac{t}{T_2})} = I_0\, e^{(-R_2 t)} \tag{5.10}$$

The ^{15}N-R_2 relaxation rate is dependent on molecular motions, as emphasized through the following equation that involves the spectral density function (Kneller *et al.*, 2002):

$$R_2 = \frac{d^2}{20}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{15}[3J(\omega_N) + 4J(0)] + R_{ex} \tag{5.11}$$

where d, c, ω_H, ω_N, γ_H and γ_N are the same as in 5.1.1.3, and R_{ex} is a slow-exchange contribution in the µs to ms time-scale that will be detailed in section 5.3.

5.1.3 Heteronuclear NOE: a cross-relaxation effect

When a spin within a molecule is at a short distance to another spin (typically less than 5 Å for two protons) its relaxation is impacted by this neighbor. In addition to the auto-relaxation processes with time-constants T_1 and T_2, a cross-relaxation effect can then be measured. It originates in the dipolar interaction between the two nuclei and translates into a variation in the polarization of the nucleus of interest.

The ^{1}H-^{15}N-NOE can be measured in a heteronuclear NOE experiment (Farrow *et al.*, 1994). Two experiments are recorded, the "saturated" experiment in which a 120° ^{1}H pulse train is applied before the standard ^{15}N-^{1}H-HSQC in order to saturate amide protons, and the "reference" experiment in which the pulse train is applied to a region where there is no proton (this provides similar conditions to the "saturated" experiment with respect to probe heating effects). The intensity of the ^{1}H-^{15}N correlation is measured in each experiment and the heteronuclear ^{1}H-^{15}N-NOE parameter is calculated:

$$NOE = \frac{I_{sat}}{I_{ref}} \tag{5.12}$$

where I_{sat} and I_{ref} are the intensities of the ^{1}H-^{15}N cross peak in the saturated and the reference experiments, respectively.

Like relaxation rates, NOE is related to molecular motions according to the following equation:

$$NOE = 1 + \frac{\gamma_H}{\gamma_N}\frac{d^2}{10}\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \times \frac{1}{R_1} \tag{5.13}$$

where d, ω_H, ω_N, γ_H and γ_N and R_1 are the same as in 5.1.1.3.

The dependence of NOE with the molecular correlation time τ_c is represented in figure 5.6. On the contrary to small molecule, the heteronuclear NOE reaches a maximum intensity and becomes field-independent for large proteins.

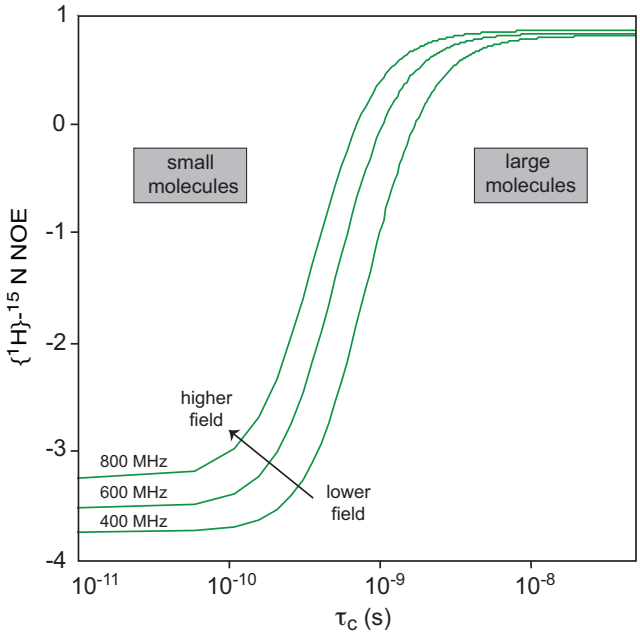


Figure 5.6: Evolution of the heteronuclear NOE with the correlation time τ_c of the molecule. Heteronuclear NOE is field-dependent for small molecules.

5.2 Relaxation and pico- to nanosecond dynamics

The fast dynamic motions can be probed by the measurement of ^{15}N relaxation rates and heteronuclear NOE. As mentioned in the introduction of the chapter, NMR relaxation measurements generally focus on ^{15}N-nuclei of amide groups. As this nuclei spin-pair can be considered as isolated, the two spins can be manipulated individually due to their large frequency difference, and the position of the ^{1}H-^{15}N-correlation is a sensitive witness for environmental changes (Kay *et al.*, 1989). ^{15}N-relaxation parameters were studied in the case of LDts to monitor backbone dynamics, nevertheless recent NMR tools have been developed to provide dynamical information on other nuclei, as ^{13}C nuclei for example, which are specifically useful to probe side-chains dynamics from methyl groups in large proteins and macromolecular assemblies (Millet *et al.*, 2002; Skrynnikov *et al.*, 2002).

5.2.1 Qualitative interpretation of ^{15}N-relaxation parameters

Once ^{15}N-relaxation experiments have been collected and relaxation parameters have been determined according to the protocol briefly described in section 5.1 for each ^{1}H-^{15}N-HSQC correlation, R_1, R_2 and heteronuclear NOE are reported as a function of residue position in the amino acid sequence (figure 5.7).

The average value of R_1 depends slightly on the molecular size (see figure 5.3). The fluctuations are related to the different orientation of the NH vector with respect to the external magnetic field B_0 and the molecular rotational diffusion tensor. On the contrary, the average value of R_2 depends drastically on the molecular size (see figure 5.3). The low values with respect to the average reveal internal motions, while

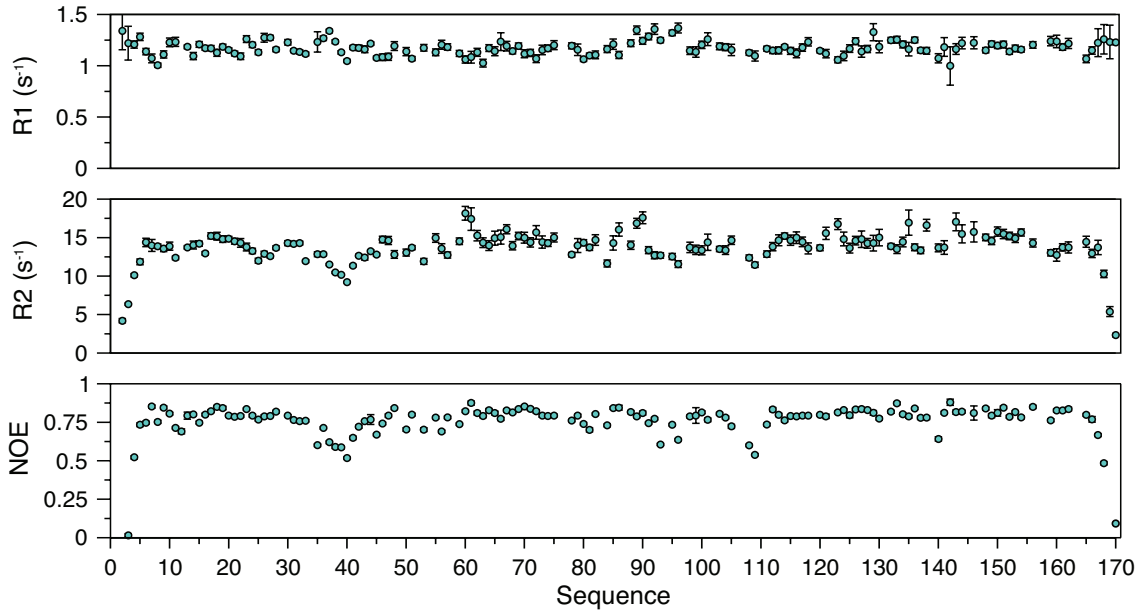


Figure 5.7: Typical representation of the R_1, R_2 and heteronuclear NOE parameters as a function of the protein sequence. The example here is taken from this thesis results on Ldt_{Bs}. The N-terminal (residues 3-7) and C-terminal (residues 168-170) regions of the protein show internal dynamics at a time-scale that is faster than the overall correlation time. Flexibility is also increased around residues 40 and 110 (hollow in the NOE and R2 values). Residues 65 and 90 tend to show chemical exchange (increase in the R_2 value).

high values indicate slow dynamics and conformational exchange. The average heteronuclear NOE value for a typical 20 kDa protein is around 0.8. Lower values point out amide vectors that undergo motions that are faster than the overall tumbling of the molecule (in the ps to ns time-scale), as for R_2. It is not rare to find N-terminal and C-terminal ends in this category as well as flexible loop regions.

Sometimes the R_2/R_1 ratio is also reported. It has the advantage to cancel or at least diminish the orientational dependence of the amide vector with respect to the magnetic field. In the absence of chemical exchange it can be used to determine the rotational correlation time τ_c for an isotropic protein (Tjandra *et al.*, 1995):

$$\frac{R_2}{R_1} \sim \frac{2}{3}\omega N^2 \tau_c^2 + \frac{7}{6} \tag{5.14}$$

On the other end, the $R_1 R_2$ product will emphasize the presence of chemical exchange, characterized by R_{ex} (Kneller *et al.*, 2002):

$$R_1 R_2 \propto \frac{S^4}{\omega N^2} + R_{ex} \tag{5.15}$$

where S is the order parameter, which is a measure of the spatial restriction of the motion. S is close to 1 if the residue is rigid and goes to zero if the residue is involved in a flexible region.

In presence of a proper structural model, R_2/R_1 ratios can be used to determine the rotational diffusion tensor by using for example the programs Modelfree (Palmer *et al.*, 1991) or TENSOR2 (Dosset *et al.*, 2000). Alternatively, in the absence of experimental data, the rotational diffusion tensor and correlation times can be calculated with HYDRONMR (García de la Torre *et al.*, 2000) and can be used to identify protein multimerization or aggregation.

5.2.2 Quantitative analysis of the ^{15}N-relaxation parameters: the Lipari-Szabo model

In the case of a spherical molecule (figure 5.8a), we saw that the spectral density function was only dependent on the correlation time of the protein τ_c (equation 5.5 on page 127) since the orientation of the amide vectors are equivalent.

However, proteins can rarely be considered as spherical molecules; instead, they can be described as ellipsoids with an anisotropic diffusion tensor (figure 5.8b). In such anisotropic conditions, the motions of the amide vectors depend on their orientation with respect to the diffusion tensor. Furthermore, the internal motions of the amide vectors must also be considered. To account for these two phenomena, a simplified approach must be used. One of the most popular methods for analyzing NMR relaxation data of proteins is the "model-free" approach, introduced by Lipari and Szabo (Lipari and Szabo, 1982a,b).

5.2.2.1 The different Lipari-Szabo models

The Lipari-Szabo (LS) model-free approach proposes to extract order parameters and time-scales of the motions from the experimentally determined R_1, R_2 and heteronuclear NOE parameters, and is based on the assumption that overall rotational tumbling of the protein and local motion of the N-H vector are completely independent. The five different models that can be used in a Lipari-Szabo (LS) analysis are summarized in Table 5.1 and schematically represented in figure 5.8.

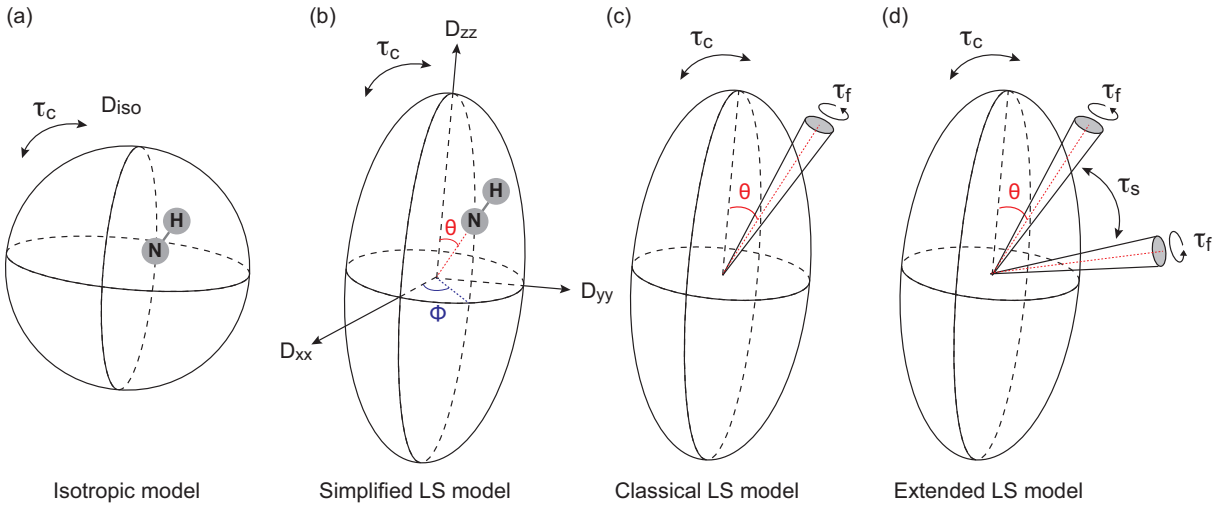


Figure 5.8: Isotropic and anisotropic molecules and their respective diffusion tensor, and schematic representation of the various Lipari-Szabo models. **a)** The molecule is a sphere and has an isotropic rotational diffusion with a coefficient D_{iso}. In this case τ_c can be calculated according to equation 5.6 (page 127). **b)** The molecule is completely anisotropic. Rotational diffusion can be described by a tensor with three principal components D_{xx}, D_{yy}, and D_{zz} along the principal axes x, y, and z , respectively. The orientation of the amide bond in the diffusion referential is given by the angles θ and ϕ. In this case, τ_c can be calculated as $\frac{1}{6D_r}$, where $D_r = 1/3(D_{xx} + D_{yy} + D_{zz})$. In this simplified Lipari-Szabo model, the spectral density function depends only on the global motion of the protein, characterized by the correlation time τ_c. **c)** Classical Lipari-Szabo model, characterized by the global rotational diffusion with correlation time (τ_c), and by one fast internal motion, represented here by a diffusion in cone (τ_f). $(1 - S^2)$ represents the amplitude of the internal motion. **d)** Extended Lipari-Szabo model for wich the spectral density function depends on three motions: the overall tumbling of the protein (τ_c), and two internal motions: one fast (τ_f) and one slower (τ_s) that can represent for example a two-site jump. The definition of the spectral density function is detailed in Table 5.1 for each model.

In the simplified Lipari-Szabo model, the motion of the amide vector is only dependent on the overall tumbling of the protein, characterized by the correlation time τ_c (figure 5.8b, Models 1 and 3 in Table 5.1).

The classical Lipari-Szabo model consists in a "two-steps" model, reducing the motion to two independent movements with different time-scales: the overall rotational tumbling of the protein and the local internal motion of the N-H vector. This model is valid only if the time-scale of fast internal motions, characterized by τ_f (in the order of picoseconds), is largely inferior to the time-scale of global motion of the protein (τ_c, in the order of nanoseconds for a protein). To get a representation of this model one could picture a diffusion in a cone where τ_f is the characteristic time-scale of the movement and $(1 - S^2)$ is the width of the cone (figure 5.8c, Models 2 and 4 in Table 5.1).

The extended Lipari-Szabo model takes into account two distinct time-scales for the description of internal motions (figure 5.8d), with two correlation times τ_f and τ_s, and two order parameters S_f^2 and S_s^2, for the $fast$ and $slow$ internal motions, respectively. When the fast internal motions are very fast (picosecond), the term $\tau_f' = \tau_c\tau_f/(\tau_c + \tau_f)$ can be neglected, leading to the spectral density function described in Model 5 in Table 5.1.

	Model	$J(\omega)$	**Parameters**
1	Simplified Lipari-Szabo, with $\tau_f << 20$ ps	$\frac{2}{5}\left[S^2\frac{\tau_c}{1+(\omega\tau_c)^2}\right]$	S^2
2	Classical Lipari-Szabo	$\frac{2}{5}\left[S^2\frac{\tau_c}{1+(\omega\tau_c)^2} + (1-S^2)\frac{\tau_e}{1+(\omega\tau_e)^2}\right]$	$S^2, \tau_e = \frac{\tau_c\tau_f}{\tau_c+\tau_f}$
3	Simplified Lipari-Szabo, including R_{ex} contribution	$\frac{2}{5}\left[S^2\frac{\tau_c}{1+(\omega\tau_c)^2}\right]$	S^2, R_{ex}
4	Classical Lipari-Szabo, including R_{ex} contribution	$\frac{2}{5}\left[S^2\frac{\tau_c}{1+(\omega\tau_c)^2} + (1-S^2)\frac{\tau_e}{1+(\omega\tau_e)^2}\right]$	S^2, τ_e, R_{ex}
5	Extended Lipari-Szabo, with $\tau_f \sim 0$ and $\tau_s \sim$ sub - ns	$\frac{2}{5}S_f^2\left[S_s^2\frac{\tau_c}{1+(\omega\tau_c)^2} + (1-S_s^2)\frac{\tau_s'}{1+(\omega\tau_s')^2}\right]$	$S_s^2, S_f^2, \tau_s' = \frac{\tau_c\tau_s}{\tau_c+\tau_s}$

Table 5.1: The 5 different models that can be used in a model-free analysis of relaxation data.

Before using one of these models for the interpretation of the relaxation parameters, it is necessary to determine the correlation time τ_c of the protein in order to remove one variable of the equations.

5.2.2.2 Determination of the correlation time τ_c and the rotational diffusion tensor

The programs the most commonly used to determine the rotational diffusion tensor and the correlation time of the protein (τ_c) are the FAST-ModelFree (for Facile Analysis and Statistical Testing for ModelFree) (Cole and Loria, 2003) program that was implemented from Modelfree and TENSOR2 (Dosset *et al.*, 2000). For LDts, we used TENSOR2, yet the two protocols are very similar as both programs are using the models based on the Lipari-Szabo approach.

Input data consists in the R_1, R_2 and NOE values for each amide, as determined from relaxation measurements, with their associated errors, and the PDB coordinates of an accurate 3D structure. The magnetic field of the spectrometer used for the relaxation experiments must also be notified. The flexible residues (*i.e.* with heteronuclear NOE greater than 0.65 and R_2 and R_1 values within 1.5 times the standard deviation to the average value over the protein) and residues presenting significant R_{ex} contribution must be first discarded from the analysis. The remaining residues are generally involved in secondary structure elements, and thus present negligible internal motions. A calculation of τ_c in the isotopic tumbling model is performed based on the average R_2/R_1 ratio of the selected residues. The correlation time is calculated from the diffusion tensor as $1/6D_{iso}$ in the isotropic case. If the isotropic model is not valid, an anisotropic model must be run and the correlation time is calculated from the diffusion tensor as $1/2(D_{xx}+D_{yy}+D_{zz})$. An anisotropic model has been used for $\mathrm{Ldt_{Bs}}$ and an isotropic model for $\mathrm{Ldt_{fm}}$.

5.2.2.3 Determination of internal mobility from relaxation parameters

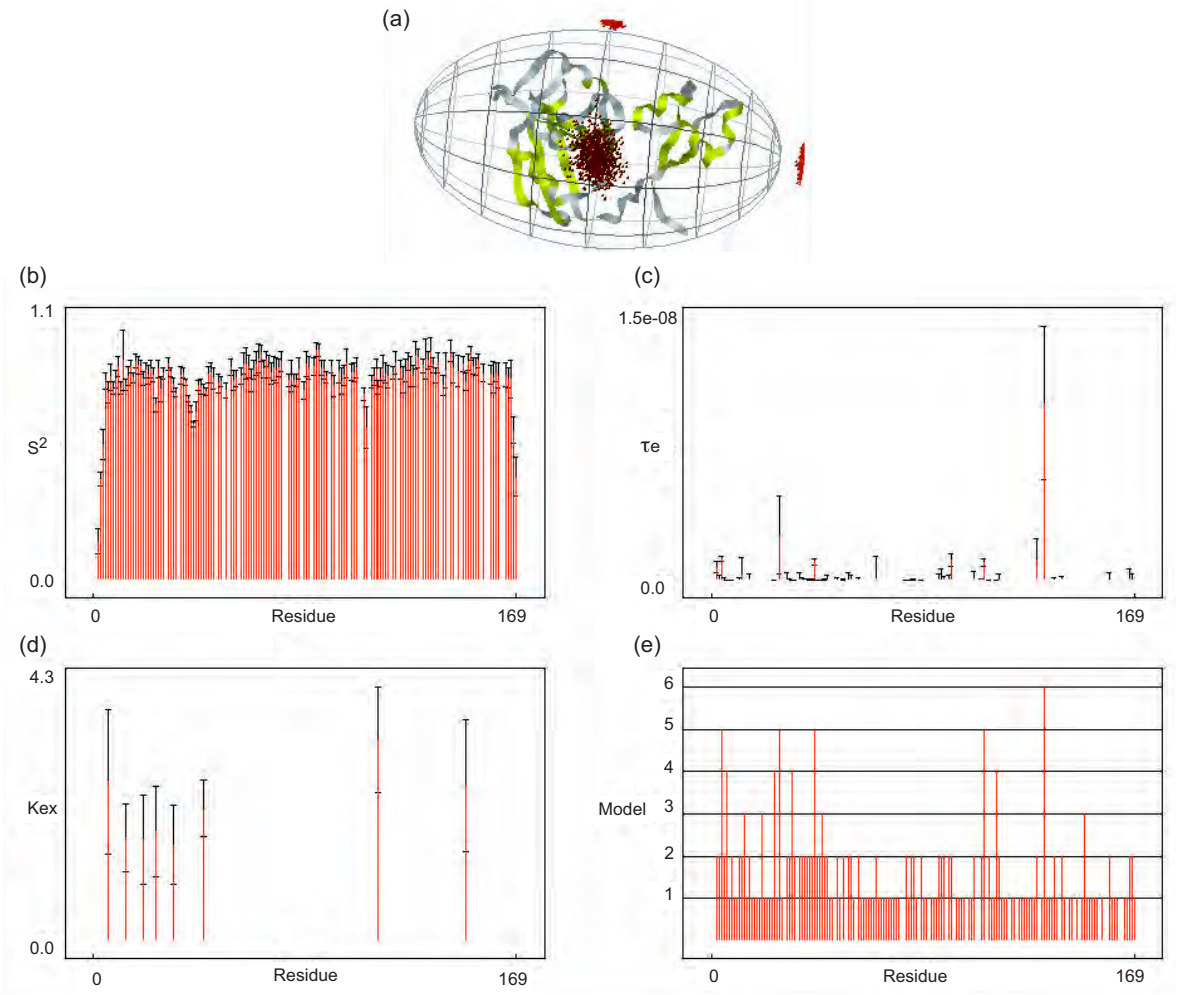


Figure 5.9: Model-free analysis of the internal mobility of $\mathrm{Ldt_{Bs}}$ using TENSOR2. a) Structure of the protein and orientation in the principal axis frame of the diffusion tensor. The four graphs depict **b)** the order parameters S^2 obtained for each ^{15}N-^{1}H vector for which the three relaxation parameters were available, **c)** the internal mobility τ_e, **d)** the exchange contribution, and **e)** the Lipari-Szabo model used for each residue. The number 6 is given when data can not satisfactorily be fitted with models 4 or 5.

Once the correlation time of the protein is known, the five Lipari-Szabo models presented in Table 5.1 are iteratively tested in TENSOR2 to determine which model fits the best the relaxation experimental data for each ^{15}N-^{1}H vector for which a set of relaxation parameters is available. Residues in conformational exchange will normally exhibit a R_{ex} contribution, while residues in the rigid part of the protein will be fitted with model 1, and residues with internal mobility will exhibit a τ_e contribution. An example of final results obtained for $\mathrm{Ldt_{Bs}}$ are shown in figure 5.9.

5.3 Relaxation and micro- to millisecond dynamics

During the analysis of fast dynamics in the ps to ns time-scales, R_{ex} contributions can be detected embedded in R_2. These contributions correspond to collective motions, also qualified as conformational exchange, and occur at the micro- to millisecond time scale. Dynamic processes on this time scales are important for the function of proteins, including catalysis, allostery, ligand binding and protein folding. Two experimental methods are utilized extensively to further characterize conformational exchange events in proteins by NMR: the Carr Purcell Meiboom Gill (CPMG) relaxation dispersion method (Carr and Purcell, 1954; Meiboom and Gill, 1958), that has been extensively modified for applications to biomolecules (Loria *et al.*, 1999a,b; Palmer, 2001), and the off-resonance R1ρ experiment.

5.3.1 Definition of conformational exchange

NMR experiments that probe conformational exchange in fact measure the excess of transverse relaxation $R_{ex} = R_2 - R_2^0$, where R_2^0 is the relaxation rate constant in absence of conformational exchange. This excess of transverse relaxation results from the exchange between different conformations or states with different chemical shifts. If an exchange process occurs between two conformations A and B with resonance frequencies ω_A and ω_B, the following equilibrium takes place:

$$A \underset{k_{BA}}{\overset{k_{AB}}{\longleftrightarrow}} B \tag{5.16}$$

where k_{AB} and k_{BA} are the kinetic constants to go from A to B and B to A, respectively. The exchange contribution to the R_2 relaxation, R_{ex}, can then be calculated as:

$$R_{ex} = \frac{p_A p_B \Delta\omega^2}{k_{ex}} \tag{5.17}$$

where k_{ex} is the exchange rate constant, $k_{ex} = k_{AB} + k_{BA}$; $\Delta\omega$ is the difference of chemical shift between the two exchanging sites A and B, $\Delta\omega = \omega_A - \omega_B$; and p_A and p_B are the populations of site A and B respectively, $p_A + p_B = 1$, $p_A = k_{BA}/k_{ex}$ and $p_B = k_{AB}/k_{ex}$.

The effect of the exchange process on the resulting NMR spectrum depends on the relative contributions of k_{ex} and $\Delta\omega$. Three different regimes can be distinguished, as depicted in figure 5.10:

- when the exchange rate is much slower than the chemical shift difference ($k_{ex} < \Delta\omega$), two distinct resonances are observed, corresponding to slow exchange. The presence of two peaks can be explained by the fact that a part of the molecules will be in conformation A and will give rise to

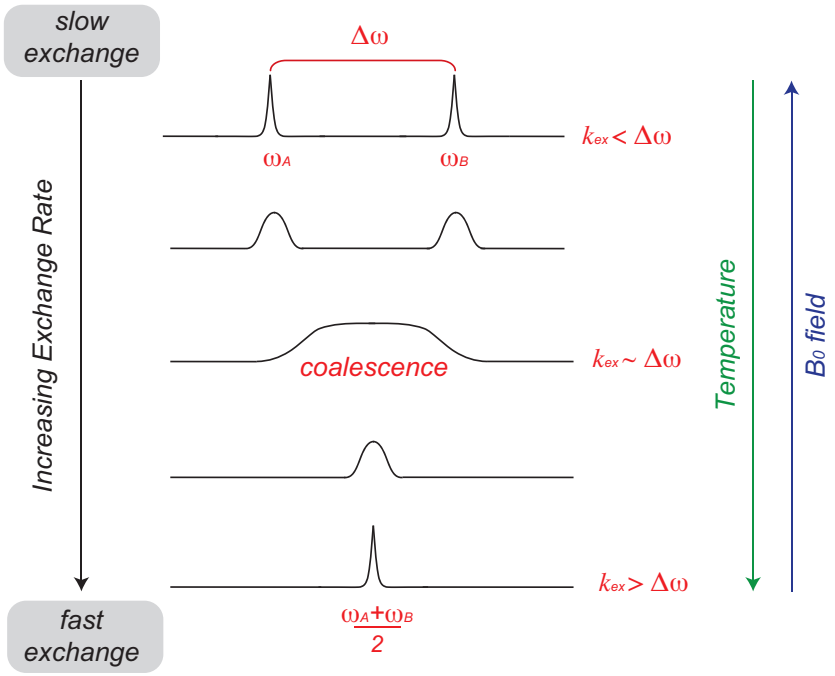


Figure 5.10: Impact of the chemical exchange regime on the appearance of the NMR signal. When $k_{ex} < \Delta\omega$, the two conformations are in slow exchange, regime characterized by two distinct peaks. When the exchange rate increases, the resonance line-width broadens, until the intermediate exchange regime, also known as *coalescence* phenomenon. Inversely, in the fast exchange regime ($k_{ex} > \Delta\omega$) a single resonance is observed, corresponding to the average position from the two individual conformations. The arrows on the right show the impact of experimental parameters (temperature and spectrometer field) on the chemical exchange regime.

a peak at chemical shift ω_A while the other part will be in conformation B and will appear at ω_B. Relative peak intensities reflect the population distribution between the two conformations.

- when the exchange rate is much faster than the chemical shift difference ($k_{ex} > \Delta\omega$), the two conformations are in fast exchange, yielding to a single peak at population-averaged position. This single peak is the result of the continuous exchange between conformations A and B at the time scale of the NMR experiment.

- when the exchange rate is in the order of the chemical shift difference ($k_{ex} \sim \Delta\omega$), intermediate exchange is observed and results in a very broad coalesced peak, and sometimes in signal disappearance if the resonance becomes indistinguishable from the baseline noise (Cavanagh *et al.*, 1995).

The relative time-scale between k_{ex} and $\Delta\omega$ can be slightly modified by different parameters, such as temperature and the magnetic field B_0. Increasing the magnetic field will favor the slow exchange regime, while increasing the temperature will favor the fast exchange regime (figure 5.10). Both parameters can thus be used to shift the equilibrium towards one or the other regime, enabling to improve quality of NMR spectra when conformational exchange occurs. This can be particularly useful to avoid intermediate exchange. Different experiments have been described to extract the populations of each state, the exchange rate and the chemical shift difference between the two states. This interpretation is much more difficult if the exchange process occurs between a larger number of states.

5.3.2 $R_{1\rho}$ experiment

In the $R_{1,\rho}$ experiment (Peng and Wagner, 1991), the magnetization is spin-locked in the rotating frame by application of a radio-frequency (rf) field. $R_{1,\rho}$ is the relaxation rate constant for the component of the magnetization along the direction of effective field in the rotating frame. $R_{1,\rho}$ depends on the amplitude of the applied rf field ω_1 and the population-averaged chemical shift Ω ($\Omega = \frac{\omega_A+\omega_B}{2}$ in the case of a two-site exchange process). The effective field in the rotating frame is defined as: $\omega_{eff} = \sqrt{(\Omega^2 + \omega_1^2)}$. During the experiment, the magnetization is aligned along this effective field that makes an angle θ with the static magnetic field B_0, $\tan\theta = \omega_1/\Omega$.

$R_{1\rho}$ rate is related to longitudinal and transverse relaxation rates by the following equation:

$$R_{1,\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \tag{5.18}$$

$R_{1,\rho}$ experiments are sensitive to conformational exchange processes if values of ω_{eff} near k_{ex} are achievable experimentally. To extract R_{2ex} contributions, $R_{1,\rho}$ measurements on biological macromolecules are performed using on-resonance or off-resonance rf fields and a relaxation-dispersion curve is obtained by varying ω_{eff}, through variations of ω_1. In the on-resonance experiment, the rf transmitter frequency is positioned close to the resonances of interest (or in the middle of the NMR spectrum) and the minimum value of ω_1 is large enough to ensure that $\theta > 70°$ over the spectral region of interest. In the off-resonance experiment, the rf transmitter is positioned far enough off-resonance to ensure that $\theta < 70°$.

5.3.3 CPMG experiment

The more commonly employed approach to characterize conformational exchange involves measuring the R_2 rate constant as a function of the frequency ν_{cpmg} of the ^{15}N 180° pulses used to refocus the ^{15}N magnetization in the CPMG spin-echo period used for Nx,y T_2 evolution. The CPMG frequency is given by:

$$\nu_{\mathrm{cpmg}} = \frac{1}{2(2\tau_{cp} + t_{180})} \tag{5.19}$$

where t_{180} is the duration of the ^{15}N-π pulse and τ_{cp} is the delay between successive π-pulses in the CPMG spin-echo train.

The CPMG spin-echo period is used to refocus the Nx,y magnetization while letting it decay with the relaxation rate constant R_2 over the whole period. In the relaxation dispersion experiment, the number of 180° pulses is varied by changing the inter-pulse delay τ_{cp}. If τ_{cp} is short (large ν_{cpmg}), the exchange between conformations A and B during τ_{cp} can not occur and the exchange contribution is small. On the contrary, if τ_{cp} is long (small ν_{cpmg}), the exchange between conformations A and B takes place during τ_{cp} and the exchange contribution can be detected, increasing the R2 value. Varying the delay τ_{cp}, thus permits to extract the R_{2ex} contribution as detailed in the following section.

It is important to note that sample heating at high pulsing rates is a major experimental limitation for this experiment. For that matter, in the laboratory it is preferably performed on the spectrometer that is equipped with a non-cryogenic probe, as cryogenic probes are more sensitive to these high pulsing rates.

5.3.4 Extraction of conformational exchange contribution

The dependence of the transverse relaxation rate on the strength of the refocusing radio-frequency field, ω_{eff} in the case of $R_{1\rho}$ and $(1/\tau_{cp})$ for CPMG experiment, defines a relaxation dispersion profile. This profile can be measured at different static magnetic fields and temperatures. Curve fitting will then enable to extract exchange rate k_{ex}, p_A and p_B populations, and chemical shift differences $\Delta\omega$.

The exchange contribution on the ^{15}N transverse relaxation correlation, $R_{2,ex}$, is calculated from CPMG experiment by:

$$R_{2,ex}(\nu_{\text{cpmg}}) = -\frac{1}{T_{relax}} \ln\left(1 - \frac{I(\nu_{\text{CPMG}})}{I_0}\right) \tag{5.20}$$

where I_0 is the cross-peak intensity in the spectrum collected at T_{relax}=0, $I(\nu_{\text{cpmg}})$ is the cross-peak intensity for a particular CPMG frequency, and T_{relax} is the constant relaxation delay used for the CPMG period (usually between 30 and 40 ms). In the case of Ldt_{Bs}, I_0 was replaced by the intensity at large ν_{cpmg} frequency (ν_{cpmg} = 1,000 Hz) where exchange contribution is negligible.

For each residue, the CPMG relaxation dispersion trajectories are obtained by plotting $R_{2,ex}$ as a function of ν_{cpmg} or τ_{cp}. Curves corresponding to residues with no exchange contribution result in a flat behavior, as the peak intensity does not depend on the number of π pulses applied, while residues involved in conformational exchange exhibit a significant decay in the relative intensity when ν_{cpmg} increases. An example of such curves for residue I^{84} in the apo- and acylated Ldt_{Bs} enzyme is given in figure 5.11.

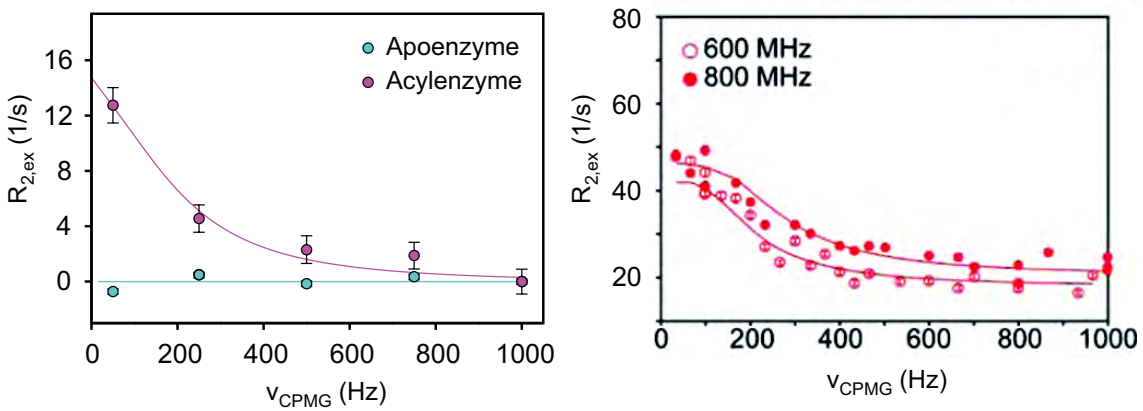


Figure 5.11: ^{15}N-CPMG relaxation dispersion trajectories. **a)** Data taken from residue I^{84} of Ldt_{Bs}, in the apoform (blue) and acylated by imipenem (magenta). $R_{2,ex}$ is calculated using equation 5.20. This residue encounters conformational exchange only upon acylation. **b)** Example taken from (Mukherjee *et al.*, 2011), showing the field dependence of the $R_{2,ex}$ for a residue in conformational exchange. In both cases, the curves are fitted with equation 5.22 for a two-site exchange model to extract the conformational exchange kinetic parameter k_{ex}.

Based on the relaxation-dispersion profiles, residues can be divided in two categories: residues with a $R_{2,ex}$ greater than a user-defined value, typically 3 s^{-1} for low ν_{cpmg}, are classified as exhibiting significant chemical exchange, while those with $R_{2,ex} < 3\ s^{-1}$ were assumed to display negligible slow motions. Residues belonging to the first category are subsequently fitted individually (or collectively if

there is an indication of a collective motion from external data or individual fits) in the frame of a two-site exchange model (Palmer, 2004):

$$R_{2,ex}(1/\tau_{cp}) = \frac{\Phi_{ex}}{k_{ex}}\left[1 - \frac{2\tanh\left(\frac{k_{ex}\tau_{cp}}{2}\right)}{k_{ex}\tau_{cp}}\right] \tag{5.21}$$

or as a function of ν_{cpmg}:

$$R_{2,ex}(\nu_{\mathrm{CPMG}}) = \frac{\Phi_{ex}}{k_{ex}}\left[1 - \frac{2\tanh\left(\frac{k_{ex}}{4}\left(\frac{1}{2\nu_{\mathrm{CPMG}}} - t_{180}\right)\right)}{\frac{k_{ex}}{2}\left(\frac{1}{2\nu_{\mathrm{CPMG}}} - t_{180}\right)}\right] \tag{5.22}$$

in which $\Phi_{ex} = p_A p_B \Delta\omega^2$. If the CPMG experiment is recorded at two different fields at least, the number of experimental data are sufficient to extract the exchange rate between the two conformers (k_{ex}), the population of the conformers (p_A and p_B), and the absolute value of the residue-specific chemical shift difference between the two states ($\mid \Delta\omega \mid$). However, if the data are only available at one field, only k_{ex} and Φ_{ex} can be extracted from the fit, but the populations and the chemical shift differences can not be explicited. The extraction of k_{ex} was possible for the acylated form of $\mathrm{Ldt_{Bs}}$ using this approach.

5.4 From protein dynamics to thermodynamics

Ligand binding often induces dynamics, either by increasing the mobility of backbone and side-chains, or inversely by reducing their mobility in or near the binding site. A decrease in the dynamics upon ligand binding is often associated with "enthalpy-entropy compensation" and "induced fit" binding, whereas an increase in the dynamics usually comes from conformational entropy that can contribute to the stabilization of complexes (Stone, 2001). Enthalpic interactions typically include hydrogen bonds, van der Waals interactions between hydrophobic side-chains and electrostatic interactions, whereas entropy describes the conformational freedom/restriction of the system (Spyracopoulos and Sykes, 2001). Relaxation is therefore a tool of choice to characterize dynamics at different time scales of an enzyme both in its apo-form and in interaction with its inhibitor, as there is a correlation between dynamical parameters derived from NMR relaxation rates and the thermodynamic entropic parameters.

The order parameter S^2 is indicative of the amplitude of spatial fluctuations experienced by a bond vector. Thus, changes in S^2 upon ligand binding or conformational transition in an enzyme reflect altered flexibility and consequent changes in conformational entropy. The motional model that is often used to interpret NMR order parameters in terms of entropy is the "diffusion-in-a-cone" model, which corresponds to diffusion in a square well potential. Using this model, the change in conformational entropy (ΔS_{conf}) can be calculated for a two-state transition:

$$\Delta S_{conf} = -k\sum_i \ln\left(\frac{3 - \sqrt{1 + 8S_{F,i}}}{3 - \sqrt{1 + 8S_{B,i}}}\right) \tag{5.23}$$

where k is Boltzmann's constant and $S_{F,i}$ and $S_{B,i}$ are the square roots of the model-free order parameters S^2 in free and bound states, respectively. From the temperature dependence of the order parameter, the heat capacity contribution from ps-ns motions can be estimated:

$$C_{p,conf} = \frac{dS_{conf}(T)}{d(\ln T)} \tag{5.24}$$

Recent studies have shown that information on conformational entropy is essential for a deep understanding of protein-ligand recognition, and should therefore be taken into account for drug design (Diehl *et al.*, 2010). NMR relaxation can be used in combination with other techniques, such as calorimetry and molecular dynamics methods, that would enable to quantitatively investigate configurational entropy contributions at an atomic resolution.

Résumé de la partie II en français

Dans cette partie sont présentés les différents outils que propose la RMN pour répondre à des questions biologiques, que ce soit au niveau structural, avec la résolution de structures d'enzymes ou de complexes enzyme-inhibiteur, thermodynamique avec la détermination de constantes d'affinité et d'ionisation, et enfin au niveau dynamique avec les études de relaxation et d'échange conformationnel. Ces trois aspects sont complémentaires, puisque c'est lors du rassemblement et de la convergence de toutes ces informations que l'on peut aller vers une compréhension des mécanismes biologiques.

Lors d'une étude structurale par RMN, le marquage isotopique de l'échantillon est requis, afin de pouvoir collecter des spectres RMN avec une sensibilité suffisante. L'obtention d'une structure passe tout d'abord par l'attribution des résonances, correspondant aux différents atomes de la protéine, du squelette protéique ainsi que des chaînes latérales. Ensuite, des contraintes structurales doivent être extraites, telles que des distances internucléaires entre protons (NOEs), et des contraintes angulaires (angles à partir des RDCs et angles de torsion). Toutes ces données peuvent alors être intégrées dans un logiciel de calcul de structure, accompagné de spectres NOESY contenant des informations sur tous les protons situés à moins de 5 Å les uns des autres. Ces spectres sont analysés et attribués de manière concomitante avec le calcul de structure, puisque les structures obtenues à chaque itération doivent être en accord avec les attributions des NOEs, qui sont elles-mêmes modifiées et complétées tout au long du processus en s'aidant du repliement de la structure. Lorsque le calcul converge vers un ensemble de structures en accord avec les données expérimentales, cet ensemble de structures calculées dans le vide est affiné dans l'eau, puis validé avant déposition dans la PDB. La RMN est également une méthode performante pour l'étude structurale de complexes protéine-ligand, notamment avec la méthode du "chemical shift mapping", qui consiste à identifier les résidus affectés par une interaction sur la base des modifications de leur déplacement chimique. Dans le cas de complexes de faible affinité, les résidus affectés peuvent ensuite être reportés sur la structure, ce qui permet d'obtenir des informations sur la localisation de l'interaction, et ceci de manière très rapide. Dans le cas de complexes de plus forte affinité, la RMN permet également par des expériences de "NOESY-filtrées", d'extraire des contraintes de distances entre une enzyme et son inhibiteur. Cette technique utilise des filtres basés soit sur la taille des molécules, soit sur le marquage isotopique dans le cas où l'un des deux partenaires seulement est marqué isotopiquement. Une fois l'interaction localisée, combinée à l'obtention de NOEs intermoléculaires, il est possible d'obtenir la structure d'un complexe enzyme/inhibiteur. Cette approche a été utilisée pour l'étude des LDts avec les carbapénèmes.

L'obtention d'une structure tridimensionnelle n'apporte malheureusement pas toutes les réponses, et l'analyse en profondeur des constantes thermodynamiques est essentielle, d'une part pour caractériser la force de l'interaction dans l'étude des complexes enzyme-inhibiteur, et d'autre part pour définir les états de protonation des résidus ionisables (Asp, Glu, Lys, Arg, His, Cys, Tyr), souvent impliqués dans les mécanismes catalytiques. Les déplacements chimiques des noyaux observés par RMN étant très sensibles au moindre changement dans leur environnement, l'analyse des perturbations de ces déplacements chimiques lors de la fixation d'un ligand ou d'un changement d'état de protonation permet la détermination de constantes de dissociation K_D et d'acidité Ka, respectivement. Connaître les valeurs des pKa des résidus ionisables dans une protéine est essentiel, puisqu'elles reflètent les rôles de ces acides aminés

dans la fonction protéique. Ainsi, des $pKas$ inhabituels vont être les témoins de l'implication dans un réseau de liaisons hydrogène ou dans des interactions électrostatiques. Combinées à la structure, ces informations vont permettre d'affiner le site actif et d'aller vers une meilleure compréhension des mécanismes dans lesquelles la protéine est impliquée.

Enfin, l'un des outils les plus puissants que possède la RMN réside dans sa sensibilité à la dynamique moléculaire sur une large gamme de temps allant de la picoseconde à la seconde. La dynamique à l'échelle rapide (ps-ns) est caractérisée par les paramètres de relaxation longitudinale (R_1), de relaxation transverse (R_2), et de relaxation croisée (NOE hétéronucléaire), alors que la dynamique à une échelle de temps plus lente (µs-ms) peut être sondée grâce aux expériences de R_{1rho} et de relaxation-dispersion. Ces derniers reflètent notamment les phénomènes d'échange conformationnel entre 2 états A et B. Les phénomènes d'échange ont des manifestations spectroscopiques différentes selon la cinétique du phénomène par comparaison avec l'échelle de temps de la RMN (représentée par le déplacement chimique) : l'échange rapide, pour lequel un seul pic de corrélation est généralement observé, correspondant à la moyenne du déplacement chimique des deux entités en échange ; l'échange lent pour lequel 2 pics sont observables, correspondant aux 2 conformations A et B ; et enfin l'échange intermédiaire, difficilement observable par RMN puisque les pics deviennent alors très larges. La dynamique étant impliquée dans un grand nombre de processus biologiques, tels que la liaison d'un ligand, le repliement des protéines, ou encore la catalyse enzymatique, elle est, tout comme la thermodynamique, elle aussi complémentaire à l'étude structurale. Pour Ldt_{Bs} nous verrons dans le chapitre suivant que la flexibilité, et plus précisément l'échange conformationnel, a un rôle déterminant dans la formation de l'acylenzyme.

Ces trois chapitres, centrés sur les méthodes RMN pour la caractérisation d'un complexe d'une enzyme avec son inhibiteur, reprennent les différentes méthodes utilisées dans ce travail pour l'étude des LDts dans leur forme libre et en interaction avec des antibiotiques. La troisième partie va maintenant s'attaquer aux résultats obtenus pour les trois enzymes étudiées parmi cette classe : les LDts de *Bacillus subtilis*, de *Enterococcus faecium* et de *Mycobacterium tuberculosis*.

Part III

Liquid-state NMR studies on L,D-transpeptidases

Chapter 6

NMR studies on the L,D-transpeptidase from *Bacillus subtilis*

The first L,D-transpeptidase to be studied in the present work is the LDt from *Bacillus subtilis*. Although it is not a pathogenic bacterial strain, this enzyme is a good model to perform a structural and dynamical analysis because the size of the protein (20 kDa) is suitable for NMR studies. The sequence of the protein is indicated in figure 6.1.

```
        10         20         30         40         50         60
GRKLLTYQVK QGDTLNSIAA DFRISTAALL QANPSLQAGL TAGQSIVIPG LPDPYTIPYH

        70         80         90        100        110        120
IAVSIGAKTL TLSLNNRVMK TYPIAVGKIL TQTPTGEFYI INRQRNPGGP FGAYWLSLSK

       130        140        150        160        170
QHYGIHGTNN PASIGKAVSK GCIRMHNKDV IELASIVPNG TRVTINRGSH HHHHH
```

Figure 6.1: Primary sequence of Ldt_{Bs}. The protein consists of 175 residues, including a 6-His tag at C-terminus position. The catalytic cysteine is highlighted in red.

The theoretical pI (isoelectric point) of Ldt_{Bs} is 10.11 according to ProtParam, a tool which allows the computation of various physical and chemical parameters for proteins, available on the molecular biology server Expasy (website: http://web.expasy.org/protparam/). The protein is thus positively charged at physiological pH as it contains 17 positively charged residues (Arg and Lys) and only 6 negatively charged residues (Asp and Glu).

Structural NMR studies were done on both the apo-Ldt_{Bs} and the acylated form with an antibiotic from the carbapenem class: the imipenem. The apoenzyme NMR sample was very stable (typically over 6 months), while the acylenzyme sample was only stable for a few days (typically 3 to 5 days, consistent with what is known from a kinetic study on Ldt_{fm} where $k_{hydrolysis} < 0.37 \times 10^{-3}$ min^{-1} (Dubée *et al.*, 2012a)) due to the hydrolysis of the thioester linkage between the enzyme and the antibiotic. The addition of an excess of imipenem was therefore necessary between two 3D NMR experiments, and new samples were regularly prepared. The protein could be concentrated up to 1 mM, and all the NMR spectra were recorded in MES buffer 12.5 mM containing 150 mM NaCl at pH 6.5, with the exception of the pH titrations which were alternatively performed in 20 mM potassium phosphate buffer or in water, in the absence of salt.

6.1 Assignment of Ldt_{Bs} NMR resonances

The assignment of the ^{1}H, ^{15}N and ^{13}C resonances of Ldt_{Bs} apo- and acylenzymes gave rise to a publication in Biomolecular NMR Assignment (Lecoq *et al.*, 2012). The chemical shift lists, together with relaxation parameters R_1, R_2 and heteronuclear NOE, were deposited in the BMRB database under the accession numbers 17701 and 18037 for apo-Ldt_{Bs} and the imipenem-acylated form of Ldt_{Bs}, respectively.

Biomol NMR Assign
DOI 10.1007/s12104-012-9358-1

ARTICLE

Backbone and side-chain ^{1}H, ^{15}N and ^{13}C assignment of apo- and imipenem-acylated L,D-transpeptidase from *Bacillus subtilis*

L. Lecoq · C. Bougault · T. Kern · J.-E. Hugonnet · C. Veckerlé · O. Pessey · M. Arthur · J.-P. Simorre

Received: 23 November 2011 / Accepted: 17 January 2012

Abstract The D,D-transpeptidase activity of Penicillin Binding Proteins (PBPs) is essential to maintain cell wall integrity. PBPs catalyze the final step of the peptidoglycan synthesis by forming 4 → 3 cross-links between two peptide stems. Recently, a novel β-lactam resistance mechanism involving L,D-transpeptidases has been identified in *Enterococcus faecium* and *Mycobacterium tuberculosis*. In this resistance pathway, the classical 4 → 3 cross-links are replaced by 3 → 3 cross-links, whose formation are catalyzed by the L,D-transpeptidases. To date, only one class of the entire β-lactam family, the carbapenems, is able to inhibit the L,D-transpeptidase activity. Nevertheless, the specificity of this inactivation is still not understood. Hence, the study of this new transpeptidase family is of considerable interest in order to understand the mechanism of the L,D-transpeptidases inhibition by carbapenems. In this context, we present herein the backbone and side-chain ^{1}H, ^{15}N and ^{13}C NMR assignment of the L,D-transpeptidase from *Bacillus subtilis* (Ldt_{Bs}) in the apo and in the acylated form with a carbapenem, the imipenem.

Keywords L,D-transpeptidase · YkuD · Antibiotic · Imipenem · Peptidoglycan · NMR assignment

L. Lecoq · C. Bougault · T. Kern · O. Pessey · J.-P. Simorre (✉)
CEA, Institut de Biologie Structurale Jean-Pierre Ebel, 41, rue Jules Horowitz, 38027 Grenoble, France
e-mail: jean-pierre.simorre@ibs.fr

L. Lecoq · C. Bougault · T. Kern · O. Pessey · J.-P. Simorre
CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, 41, rue Jules Horowitz, 38027 Grenoble, France

L. Lecoq · C. Bougault · T. Kern · O. Pessey · J.-P. Simorre
Université Joseph Fourier—Grenoble 1, Institut de Biologie Structurale Jean-Pierre Ebel, 41, rue Jules Horowitz, 38027 Grenoble, France

J.-E. Hugonnet · C. Veckerlé · M. Arthur
Centre de Recherche des Cordeliers, LRMA, Université Pierre et Marie Curie—Paris 6, UMR S 872, 75006 Paris, France

J.-E. Hugonnet · C. Veckerlé · M. Arthur
Université Paris Descartes, UMR S 872, 75006 Paris, France

J.-E. Hugonnet · C. Veckerlé · M. Arthur
INSERM, U872, 75006 Paris, France

Biological context

β-lactams are one of the oldest and still most broadly used class of antibiotics for the treatment of severe bacterial infections. The D,D-transpeptidases, which catalyze the final cross-linking step of peptidoglycan synthesis in the bacterial cell wall, are the main target of β-lactams since transpeptidation is essential to maintain the integrity of the cell wall (Ghuysen and Shockman 1973). These enzymes belong to a large family of active-site serine acyl-transferases that bind penicillin covalently and are thus referred to as penicillin-binding proteins (PBPs) (Goffin and Ghuysen 1998). Until recently, D,D-transpeptidases, which are highly redundant enzymes, were considered as the essential target of β-lactam antibiotics, as the latter are structure analogues of the D-Ala-D-Ala extremity of stem peptides (Tipper and Strominger 1965) and act as suicide substrates of the D,D-transpeptidases.

A novel mechanism of β-lactam resistance has been described in a mutant of *Enterococcus faecium* (Mainardi et al. 2000, 2005), in which by-pass of D,D-transpeptidases by L,D-transpeptidases leads to high-level resistance to ampicillin. The L,D-transpeptidases form unusual peptidoglycan cross-links, and harbor a different catalytic nucleophile in the

Published online: 26 January 2012

form of an invariant cysteine residue. Unexpectedly, a specific class of β-lactams, the carbapenems, is able to inactivate the enzyme by acylation of the catalytic cysteine (Mainardi et al. 2007). As the L,D-transpeptidation is the predominant (80%) mode of peptidoglycan cross-linking in *Mycobacterium tuberculosis* (Lavollay et al. 2008), L,D-transpeptidases are promising targets for development of anti-tuberculous drugs belonging to the carbapenem class.

To date, only two L,D-transpeptidase structures have been solved, from two bacterial strains: *E. faecium* (Biarrotte-Sorin et al. 2006) and *Bacillus subtilis* (Bielnicki et al. 2006). The specificity of carbapenems against other β-lactams is still not understood, and no structure of the acylated form of L,D-transpeptidases by antibiotics is available. To understand the interaction between these enzymes and carbapenems, we report here the backbone and side-chain assignments of the 169 amino acid long *B. subtilis* L,D-transpeptidase (Ldt_{Bs}) in the apo form and in the imipenem-acylated form.

Methods and experiments

Cloning, expression and purification of Ldt_{Bs}

The open reading frame encoding Ldt_{Bs} with a hexahistidine tag in C-terminal was amplified with primers 5′-AA CCATGGGGCTGCTTACGTACCAGGTGAAGC-3′ and 5′-TTGGATCCCCGGT-TAATCGTGACTCTCGT-3′. The PCR product digested with NcoI and BamHI (underlined) was cloned into pET2818. The corresponding ^{13}C,^{15}N-uniformly labeled protein was expressed in *Escherichia coli* strain BL21(DE3) using M9 minimal media containing 2.0 g/L ^{13}C-glucose and 1.0 g/L $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) and supplemented with 0.1 mg/mL ampicillin, 1 mg/mL of pyridoxin, 1 mg/mL of biotin, 1 mg/mL of panthothenate hemi calcium, 1 mg/mL of folic acid, 1 mg/mL of cholin chloride, 1 mg/mL of niacineamide, 0.1 mg/mL of riboflavin, and 5 mg/mL of thiamin. Bacterial cultures were grown overnight at 37 °C to an optical density OD_{600} of 0.8. Expression was induced by adding 0.5 mM IPTG and incubating at 16 °C for 18 h. Cells were centrifuged. The cell pellet was resuspended in a 50 mM Tris–HCl and 300 mM NaCl buffer at pH 7.5 and disrupted with a French press. Cell debris were removed by centrifugation (48,000 g at 4 °C for 30 min) and the clarified extract was loaded onto a nickel-nitrilotriacetate agarose (Ni–NTA) resin (Qiagen, GmbH, Germany). Proteins were eluted with 500 mM imidazole (in 50 mM Tris–HCl, 300 mM NaCl, pH 7.5) and further purified onto a Superdex 75 (GE-Healthcare) gel-filtration column equilibrated with a 12.5 mM MES, 150 mM NaCl buffer at pH 6.5. The protein elution was monitored at 280 nm on a Biologic Duo Flow FPLC system (Bio-Rad). Protein purity was assessed by SDS-PAGE gel electrophoresis and mass spectrometry. The observed (calculated) molecular weight was 20,009.88 Da (20,008.70 Da) on a 6210 LC-TOF Agilent mass spectrometer.

NMR experiments

The samples were prepared in a buffer containing 12.5 mM MES and 150 mM NaCl buffer at pH 6.5 with 10% D_2O. To prepare the acyl-enzyme sample, Ldt_{Bs} was incubated with 2.5 equivalents of imipenem (gift from Merck) in the same conditions than the apo-enzyme. Data were collected on 0.7 mM ^{13}C,^{15}N-labeled protein samples for both Ldt_{Bs} apo- and acyl-enzymes. Spectra for backbone and histidine side-chain assignments were recorded at 298 K on a Varian Direct Drive spectrometer equipped with a cryogenic triple ^{1}H, ^{13}C, ^{15}N-resonance probe functioning with a ^{1}H NMR frequency at 600 MHz (Agilent). 3D ^{15}N-NOESY-HSQC (τ_m = 140 ms), and methyl-selective assignment experiments (Van Melckebeke et al. 2004) were recorded using a Varian Direct Drive spectrometer operating at 800 MHz and equipped with a cryogenic triple resonance probe (Agilent). Finally, 3D aliphatic ^{13}C-NOESY-HSQC (τ_m = 120 ms) experiment was collected in the CERM Bio-NMR facility on a Bruker Avance spectrometer operating at 900 MHz equipped with a cryogenic triple ^{1}H, ^{13}C, ^{15}N-resonance probe. Backbone sequential resonances were assigned using 3D HNCO, HNCACB, CBCA(CO)NH and side-chain resonances were assigned using aliphatic and aromatic ^{1}H–^{13}C CT-HSQC, 3D (H)CCH-TOCSY and 3D NOESY-HSQC experiments (Sattler et al. 1999). The histidine side-chains were assigned using a ^{1}H-^{15}N-HMQC experiment optimized for $^{2}J_{HN}$ coupling transfers (Pelton et al. 1993). NMR data were processed with NMRPipe (Delaglio et al. 1995) and analyzed using the CCPNMR Analysis 2.2 software (Vranken et al. 2005).

Assignments and data deposition

A 2D ^{1}H–^{15}N HSQC spectrum of assigned Ldt_{Bs} is shown in Fig. 1a, and a superposition of the apo- and acyl-enzymes spectra is provided in Fig. 1b. According to the dispersion and the resolution of these two spectra, the protein is clearly folded in both forms. The overall backbone assignments on the entire Ldt_{Bs} protein, excluding the hexahistidine tag at the C-terminus, have been completed to 100% for the non-proline $^{1}H_N$–^{15}N resonances (with the exception of N-terminal residue), 100% for the $^{13}C_\alpha$, 99.3% for the $^{13}C_\beta$ and 98.8% for the $^{13}C'$ resonances. For the acyl-enzyme, three regions are particularly affected by the antibiotic. Residues in these regions show an intensity decrease in the ^{1}H–^{15}N HSQC spectrum, leading in some cases to signal extinction (Fig. 2). This resulted in an

Backbone and side-chain ^{1}H, ^{15}N and ^{13}C assignment

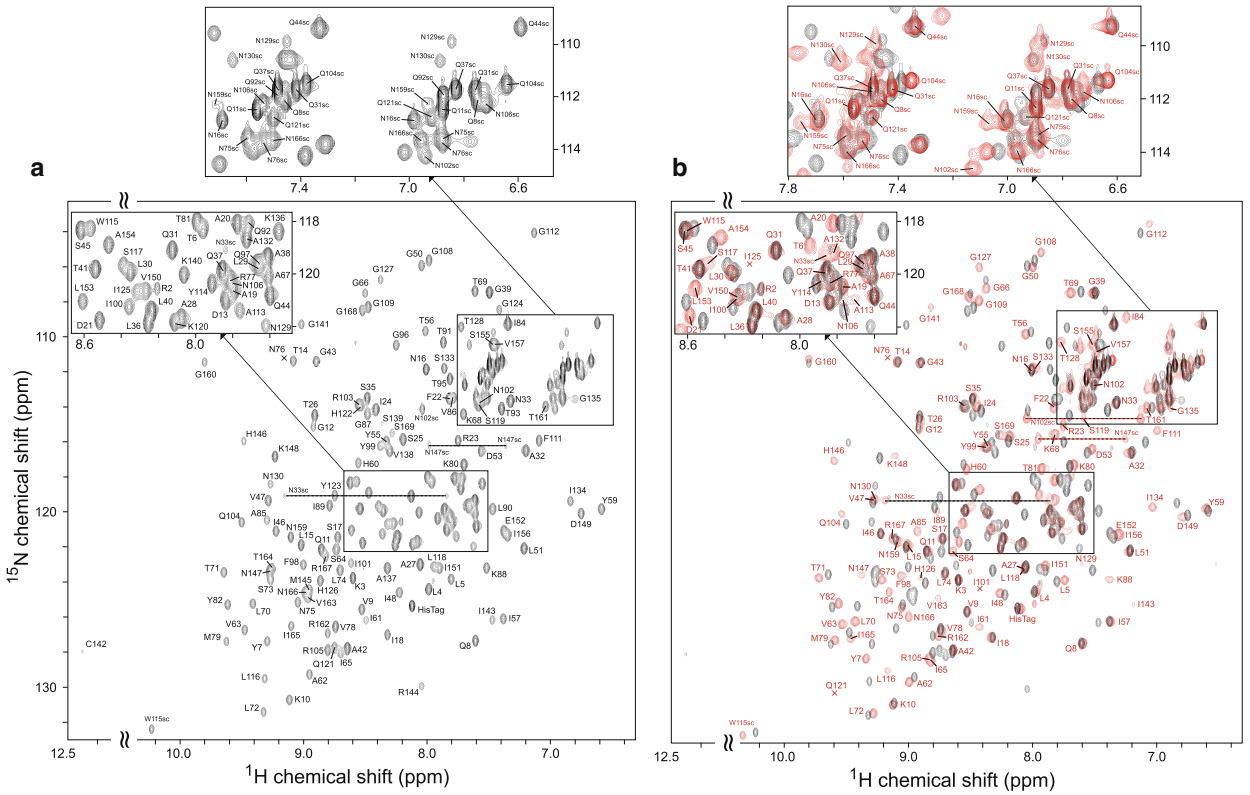


Fig. 1 **a** Fully assigned ^{1}H–^{15}N HSQC spectrum of 0.7 mM Ldt_{Bs} in 12.5 mM MES buffer, 150 mM NaCl at pH 6.5. The spectrum was recorded at 25 °C at 600 MHz ^{1}H frequency. The side-chains are labeled as 'sc' (Nε1-Hε1 for Trp and NH_2 for Gln and Asn) and the specific NH_2 region is shown above the spectrum. **b** Superposition of ^{1}H–^{15}N HSQC spectrum of 0.7 mM Ldt_{Bs} (*black*) and imipenem-acylated Ldt_{Bs} (*red*) in 12.5 mM MES buffer, 150 mM NaCl at pH 6.5. Both spectra were recorded at 25 °C at 600 MHz ^{1}H frequency. The assignment is provided for the acyl-enzyme in *red*. Crosses indicate the position of lower intensity resonances

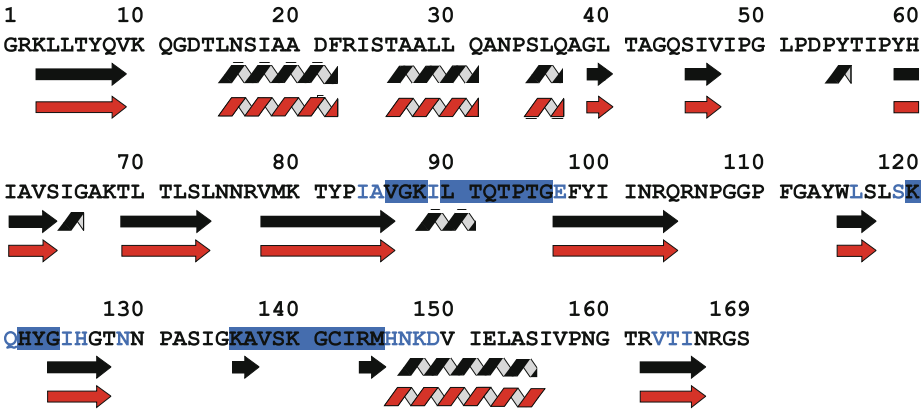


Fig. 2 Amino acid sequence with the secondary structure prediction of Ldt_{Bs} apo- (*black*) and acyl-enzyme (*red*) given by TALOS+. The missing residues in the ^{1}H–^{15}N HSQC spectrum of the acyl-enzyme are highlighted in *blue* in the sequence, and the residues with an intensity decrease larger than 50% compared to the apo-enzyme are written in *blue*. Helix and *β*-sheet secondary structures are represented by helices and *arrows*, respectively

assignment completeness of 86.0% for the non-proline $^{1}H_N$–^{15}N resonances (with the exception of the N-terminal residue), 88.7% for the $^{13}C_\alpha$, 89.5% for the $^{13}C_\beta$ and 78.0% for the $^{13}C'$ resonances. In total, 95.7% of the ^{1}H side chain resonances were assigned in the apo-enzyme, against 84.4% for the acyl-enzyme. Chemical shifts were deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession numbers 17701 for apo-Ldt_{Bs} and 18037 for the imipenem-acylated form of Ldt_{Bs}. Analysis of the chemical shifts using TALOS+ (Shen et al. 2009) is summarized in Fig. 2 and shows that both forms present similar predicted secondary structure elements, except in the three regions where conformational exchange occur in the imipenem-acylated Ldt_{Bs}.

Acknowledgments Financial support by the Access to Research Infrastructures activity in the 6th Framework Programme of the EC (Contract # RII3-026145, EU-NMR) for conducting the research on the 900 MHz spectrometer of the CERM Facility and by the French TGIR-RMN for conducting the research on the 800 MHz spectrometer of the IBS Facility are gratefully acknowledged.

References

Biarrotte-Sorin S et al (2006) Crystal structure of a novel beta-lactam-insensitive peptidoglycan transpeptidase. J Mol Biol 359:533–538

Bielnicki J et al (2006) *B. subtilis* YkuD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. Proteins 62:144–151

Delaglio F et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Ghuysen JM, Shockman GD (1973) Bacterial membranes and wall. In: Lièvе L (ed) Marcel Dekker. New York, NY, pp 37–130

Goffin C, Ghuysen JM (1998) Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. Microbiol Mol Biol Rev 62:1079–1093

Lavollay M et al (2008) The peptidoglycan of stationary-phase mycobacterium tuberculosis predominantly contains cross-links generated by L,D-transpeptidation. J Bacteriol 190:4360–4366

Mainardi JL et al (2000) Novel mechanism of beta-lactam resistance due to bypass of D,D-transpeptidation in *Enterococcus faecium*. J Biol Chem 275:16490–16496

Mainardi JL et al (2005) A novel peptidoglycan cross-linking enzyme for a beta-lactam-resistant transpeptidation pathway. J Biol Chem 280:38146–38152

Mainardi JL et al (2007) Unexpected inhibition of peptidoglycan L,D-transpeptidase from *Enterococcus faecium* by the beta-lactam imipenem. J Biol Chem 282:30414–30422

Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated III(Glc), a signal-transducing protein from *Escherichia coli*, using 2-dimensional heteronuclear NMR techniques. Protein Sci 2:543–558

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223

Tipper DJ, Strominger JL (1965) Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. Proc Natl Acad Sci USA 54:1133–1141

Van Melckebeke H, Simorre JP, Brutscher B (2004) Aminoacid-type edited NMR experiments for methyl–methyl distance measurement in ^{13}C-labeled proteins. J Am Chem Soc 126:9584–9591

Vranken WF, Boucher W et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

6.2 Structure and dynamics characterization of Ldt_{Bs}: A brief summary of the article published in Structure

The large panel of methods offered by NMR spectroscopy presented in the second part of the present work were used to investigate the structural, thermodynamical and dynamical properties of apo- and imipenem-bound L,D-transpeptidase from *B. subtilis* and to characterize its Cys-His-Gly catalytic triad. Thermodynamics of the non-covalent complex were also explored.

The first part of the work on Ldt_{Bs} focused on the determination of the structure of the apoenzyme. As the X-ray structure available for Ldt_{Bs} was crystallized with a cadmium ion in the active site, responsible for protein dimerization and directly coordinated to the active Cys^{142}, NMR was extensively used to reach a correct definition of the active site of the apoenzyme in an environment closer to physiological conditions. The first barrier to overcome was the determination of the protonation states of the ionizable residues involved in the catalytic triad, *i.e.* His^{126} and Cys^{142}. Chemical shift patterns of histidine residues and pH titration were used with this aim in view, following the method explained in section 4.2. His^{126} was unambiguously established as a His residue with unusually low pKa below 3.1 (which stands over 3.5 units lower than the average histidine pKa value in proteins, see Table 4.1 on page 113) and exists as the less common neutral Nδ1 tautomer over a wide pH range, from pH 3.6 to pH 10.0. The behavior of Cys^{142} was also surprising, with no changes in the Cβ and Hβ chemical shifts, revealing that its protonation state remains the same over the whole pH range and that its pKa is found outside these limits. Even if the proton attached to the sulfur atom of the cysteine was not directly visible on the NMR spectra, a careful structural analysis was performed on the final NMR structure ensemble and demonstrated that Cys^{142} remains protonated and has therefore a pKa above 10.5 (over 3.7 units higher than the average cysteine pKa value in proteins). This analysis led to the conclusion that Cys^{142} is involved in a hydrogen bond network, with the presence of a hydrogen bond between the Sγ-Hγ of the cysteine and the unprotonated Nϵ2 of His^{126}. On the other side of the histidine ring, the Nδ1-Hδ1 forms a hydrogen bond with the backbone oxygen atom of Gly^{127}. Consequently, instead of the generally accepted thiolate-imidazolium (S^-/NH^+) ion pair found in the active site of the cysteine protease papain (Lewis *et al.*, 1981), Ldt_{Bs} exploits a neutral, hydrogen-bonded thiol-imidazole pair (SH/N) for catalysis. A new mechanism for the acylation has therefore been suggested, in accordance with these experimental findings.

The second part of the study on Ldt_{Bs} was done on the non-covalent complex with different antibiotics. The mutation of Cys^{142} into Ala prevented the formation of a covalent bond between the β-lactam cycle and the sulfur atom of the cysteine. Antibiotics from two β-lactam families were tested: imipenem and ampicillin, representing the carbapenem and the penicillin family, respectively. The dissociation constants were measured and revealed similar weak affinity ($K_D > 10$ - 100 mM) for these two antibiotics, suggesting that the formation of the Michaelis complex is not the determinant step for the selectivity towards carbapenems. Furthermore the chemical shift mapping pointed out that both antibiotics interact with the enzyme within a similar binding pocket, located in the vicinity of the catalytic cysteine.

In the last part of the study, we highlighted a significant structural and dynamical rearrangement induced by the acylation of Ldt_{Bs} by imipenem. The first clue to the increase in mobility of residues near the active site of the complex came from the observed attenuation or disappearance of their NMR signals. The second and last clue came from proper dynamics measurements on the acylenzyme and on another covalent adduct through a disulfide. The ms-to-μs dynamics hampered the measurement of

NMR-derived restraints for many of these residues, as well as for the bound imipenem, thus resulting in a particularly high RMSD value for the drug and at the interaction interfaces within the final structural ensemble calculated for the acylenzyme. Relaxation-dispersion experiments, which sensitively detect μs-ms time-scale motions between exchanging conformational states with different chemical shifts, offered a dynamics description of the enzyme. In contrast to the Ldt_{Bs} apoenzyme, which displayed negligible slow motions, we demonstrated that three regions of imipenem-bound Ldt_{Bs} located at the interface with the active site undergo significant slow motions. Exchange rate constants of approximately 2 ms^{-1} were individually measured for these three regions. As the number of detectable probes in each region is very small, it is nevertheless speculative to propose a cooperative motion at this stage. To provide more clue on the origin of these motions, a 2-nitro-5-thiobenzoate (TNB) molecule was covalently attached to Cys^{142} through a disulfide linkage. Dynamics measurements performed on Ldt_{Bs}-TNB also revealed significant slow motions with exchange rate constants of approximately 6 ms^{-1}. Regions exhibiting conformational exchange upon TNB binding were exactly the same than the regions affected by imipenem binding. Furthermore, these slow motions result neither from the loss of the His^{126}-Cys^{142} hydrogen bond, because the C142A mutant did not show such dynamics, nor from the non-covalent binding, as increased flexibility was not observed in this case. Together, these results demonstrate that the induced motions are uniquely due to the covalent modification of the enzyme, whatever is the nature of the covalent binding (thioester linkage with imipenem or disulfide linkage with TNB). These induced dynamics could be catalytically relevant, perhaps facilitating the second step of the transpeptidation reaction.

6.3 Article published in Structure

The article was published in Structure in May 2012 and this work was highlighted in an article by Soumya De and Lawrence McIntosh, from the University of British Columbia, Vancouver, in the same issue.

Coordinates of the 20-structures ensemble for the apo- and imipenem-bound forms of Ldt_{Bs} were deposited in the PDB under the accession codes 3ZQD and 4A52, respectively.

Structure
Article

Dynamics Induced by β-Lactam Antibiotics in the Active Site of *Bacillus subtilis* L,D-Transpeptidase

Lauriane Lecoq,[1,2,3] Catherine Bougault,[1,2,3] Jean-Emmanuel Hugonnet,[4,5,6] Carole Veckerlé,[4,5,6] Ombeline Pessey,[1,2,3] Michel Arthur,[4,5,6,*] and Jean-Pierre Simorre[1,2,3,*]
[1]CEA, DSV, Institut de Biologie Structurale (IBS), Grenoble F-38027, France
[2]CNRS, UMR 5075, 41 Rue Jules Horowitz, Grenoble F-38027, France
[3]Université Joseph Fourier, Grenoble 1 F-38000, France
[4]Centre de Recherche des Cordeliers, LRMA, Equipe 12, Université Pierre et Marie Curie - Paris 6, UMR S 872, Paris F-75006, France
[5]Université Paris Descartes, Sorbonne Paris Cité, UMR S 872, Paris F-75006, France
[6]INSERM, U872, Paris F-75006, France
*Correspondence: michel.arthur@crc.jussieu.fr (M.A.), jean-pierre.simorre@ibs.fr (J.-P.S.)
DOI 10.1016/j.str.2012.03.015

SUMMARY

β-lactams inhibit peptidoglycan polymerization by acting as suicide substrates of essential D,D-transpeptidases. Bypass of these enzymes by unrelated L,D-transpeptidases results in β-lactam resistance, although carbapenems remain unexpectedly active. To gain insight into carbapenem specificity of L,D-transpeptidases (Ldts), we solved the nuclear magnetic resonance (NMR) structures of apo and imipenem-acylated *Bacillus subtilis* Ldt and show that the cysteine nucleophile is present as a neutral imidazole-sulfhydryl pair in the substrate-free enzyme. NMR relaxation dispersion does not reveal any preexisting conformational exchange in the apoenzyme, and change in flexibility is not observed upon noncovalent binding of β-lactams (K_D > 37.5 mM). In contrast, covalent modification of active cysteine by both carbapenems and 2-nitro-5-thiobenzoate induces backbone flexibility that does not result from disruption of the imidazole-sulfhydryl proton interaction or steric hindrance. The chemical step of the reaction determines enzyme specificity since no differences in drug affinity were observed.

INTRODUCTION

Antibiotics of the β-lactam family have been extensively used to treat bacterial infections in the past 70 years since these broad-spectrum antibiotics combine excellent bioavailability and limited toxicity. β-lactams inhibit the last polymerization step of peptidoglycan, a giant macromolecule of approximately 10^9–10^{10} Da, which surrounds the cytoplasmic membrane and acts as a stress-bearing layer to protect the bacterium from the osmotic pressure of the cytoplasm (Vollmer and Seligman, 2010). This essential role of peptidoglycan is conferred by a net-like structure consisting of glycan strands crosslinked by the formation of amide bonds between short stem peptides. Since β-lactams are structural analogs of the D-Ala-D-Ala extremity of stem peptides (Tipper and Strominger, 1965; Figure 1A), the antibiotics act as suicide substrates of the D,D-transpeptidases that catalyze the essential crosslinking step of peptidoglycan synthesis (Zapun et al., 2008; Figure 1B).

Until recently, D,D-transpeptidases, which are highly conserved in the entire bacterial kingdom, were considered as the lone essential target of β-lactam antibiotics. We have recently shown that antibiotics belonging to a particular class of β-lactams, the carbapenems (Figure 1D), inactivate a second class of crosslinking enzymes, the L,D-transpeptidases (Mainardi et al., 2005), that bypass the classical D,D-transpeptidases by forming unusual peptidoglycan crosslinks (Mainardi et al., 2000; Figures 1B and 1C). Transpeptidases with D,D- and L,D-specificities are structurally unrelated and harbor different catalytic nucleophiles in the form of invariant Ser (Sauvage et al., 2008) and Cys (Biarrotte-Sorin et al., 2006) residues that are acylated by all β-lactams (Pratt, 2008) or by carbapenems only (Mainardi et al., 2007), respectively. Genes encoding L,D-transpeptidases are sporadically distributed in bacteria and these enzymes generally have marginal and unessential roles in peptidoglycan synthesis (Magnet et al., 2007) except in a few Gram-positive bacteria (Mainardi et al., 2008), including *Mycobacterium tuberculosis* (Lavollay et al., 2008), the agent of tuberculosis, *M. abscessus* (Lavollay et al., 2011), a related human pathogen, and *Clostridium difficile* (Peltier et al., 2011), responsible for intestinal diseases associated with antibiotic therapy. In *M. tuberculosis*, L,D-transpeptidation is the predominant (80%) mode of peptidoglycan crosslinking (Lavollay et al., 2008). The five L,D-transpeptidase paralogs of this bacterium (Gupta et al., 2010) include enzymes essential for virulence, as well as those that are the likely targets of carbapenems (Lavollay et al., 2008; Figure 1D), which are currently being investigated as therapeutic alternatives in the treatment of extensively drug-resistant tuberculosis (Hugonnet et al., 2009).

Bypass of the classical D,D-transpeptidases was detected for the first time in mutants of *Enterococcus faecium* that had acquired high-level resistance to the β-lactam ampicillin in vitro (Figure 1D) and relied exclusively on an L,D-transpeptidase,

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

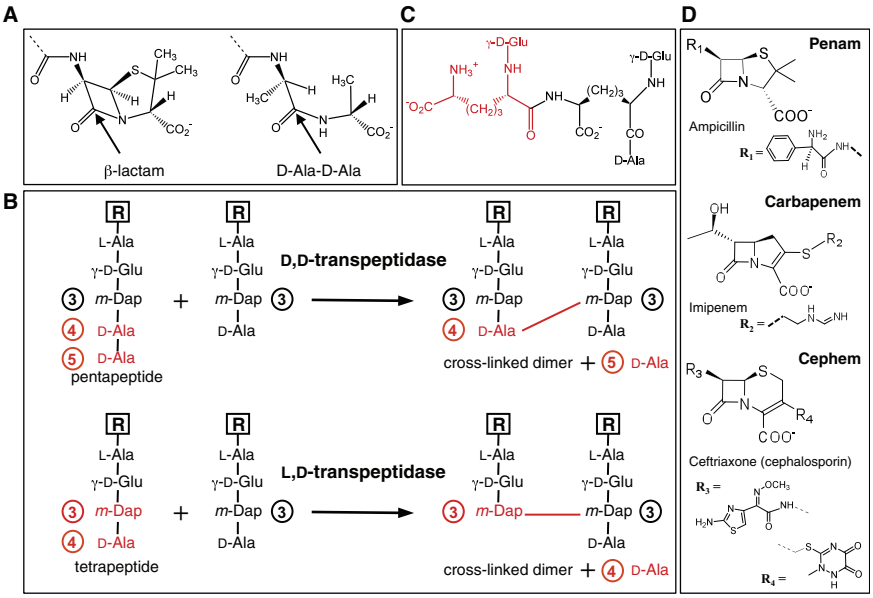


Figure 1. Mechanism of Action of β-Lactam Antibiotics

(A) Structural similarity between β-lactam antibiotics and the D-Ala-D-Ala extremity of peptidoglycan stem peptides. The arrows indicate the sites of nucleophilic attack by the D,D-transpeptidases in the suicide (β-lactam) and natural (stem peptide) substrates, respectively.

(B) Peptidoglycan crosslinking reactions catalyzed by transpeptidases with D,D- and L,D- specificities. The first step of the reaction involves the nucleophilic attack of the D-Ala^4 or *m*-Dap^3 carbonyl by the transpeptidase enzyme and leads to the formation of an ester or a thioester bond between the carbonyl and the active site Ser or Cys residue of D,D- and L,D-transpeptidases, respectively. The second step involves nucleophilic attack of the resulting acylenzyme by the side chain amine of *m*-Dap in the second stem-peptide substrate and leads to formation of 4→3 or 3→3 crosslinks. *m*-Dap, *meso*-diaminopimelic acid. R, glycan strands.

(C) Structure of the *m*-Dap-*m*-Dap 3-3 crosslink. *m*-Dap from the acyl donor and acceptor peptide strands are depicted in red and black, respectively.

(D) Chemical structures of the three main classes of β-lactams.

Ldt_{fm}, for peptidoglycan crosslinking (Mainardi et al., 2005). Inactivation of Ldt_{fm} by carbapenems was unexpected (Mainardi et al., 2007) since β-lactams are thought to be structural analogs of the D-Ala-D-Ala extremity of peptidoglycan stem pentapeptides (Tipper and Strominger, 1965), a structural motif that is absent from their substrate, as these enzymes cleave the preceding L-D peptide bond of a stem tetrapeptide (Mainardi et al., 2007).

The crystal structures of two related L,D-transpeptidases have been determined, Ldt_{fm} from *E. faecium* (Biarrotte-Sorin et al., 2006) and Ldt_{Bs} from *Bacillus subtilis* (Bielnicki et al., 2006), but these structures failed to further our understanding of the mechanism of inactivation of L,D-transpeptidases by carbapenems since the active site of the crystallized form of the enzymes was artifactually occupied by a divalent cation (Cd^{2+} or Zn^{2+}) and a bulky anion (SO_4^{2-}), which modified the position of catalytic residues. In addition, covalent adducts resulting from acylation of the active site Cys residue of Ldt_{fm} did not lead to crystals that could be exploited for collection of high-resolution diffraction data (unpublished results). Liquid-state NMR has therefore been used here to gain insight into the interaction of L,D-transpeptidases with β-lactams. We solved the solution structure of two biologically relevant forms of Ldt_{Bs}, the apoenzyme and the acylenzyme resulting from the nucleophilic attack of the imipenem β-lactam carbonyl by the catalytic cysteine. The enzyme proton shuttle mechanism has been investigated by defining the hydrogen-bonding and ionization (pKa) states of the Cys-His-Gly catalytic triad in the apoenzyme. Ldt_{Bs} acylation by imipenem induces conformational dynamics in the active site in the millisecond regime. Since the noncovalent interaction of Ldt_{Bs} with both penams and carbapenems exhibits similar low affinities, we propose a modified induced-fit mechanism for enzyme acylation in which carbapenems specificity is defined at the chemical step of the reaction.

RESULTS

NMR Structure of Ldt_{Bs}

The structure of *B. subtilis* L,D-transpeptidase (Ldt_{Bs}) fused to a C-terminal hexahistidine affinity tag (Figure 2) was determined using multidimensional NMR spectroscopy in solution, at 25°C, in a buffer containing 12.5 mM MES and 150 mM NaCl (pH 6.5) (see Figure S1 available online for the [^{1}H,^{15}N-HSQC] spectrum). Distance restraints (3,191) derived from nuclear Overhauser effect (NOE), 143 dihedral constraints derived from chemical-shift data, and 169 angular restraints derived from residual dipolar couplings (RDCs) were introduced in the final structure calculation protocol, as well as the experimentally determined protonation states of active-site residues (see below). The rmsds, calculated for the 20 lowest energy structures with respect to the average structure, were 0.31 Å for backbone

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

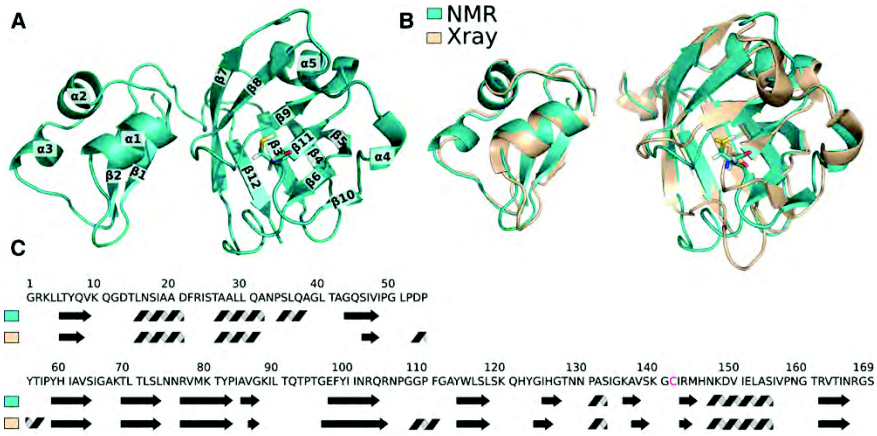


Figure 2. Structure of *B. subtilis* L,D-transpeptidase Ldt_{Bs}

(A) Lowest-energy NMR structure with emphasis on the secondary structure elements.

(B) Superposition of the NMR (blue teal) and X-ray (wheat) structures of the LysM (left) and catalytic (right) domains. Cys142 is shown as sticks.

(C) Overlay of the amino acid sequence and NMR (top) as well as X-ray (bottom) secondary structure elements. Catalytic Cys142 is highlighted in red and the C-terminal histidine tag is omitted. Numbering of the amino acids in the X-ray structure was shifted by +3 to match the NMR sequence numbering.

See also Figure S1.

atoms and 0.70 Å for all heavy atoms. These results, together with the structural statistics reported in Table 1, show that the structure is extremely well defined.

Ldt_{Bs} comprises two domains, an N-terminal LysM peptidoglycan-binding domain, from residues 1 to 54, and a C-terminal catalytic domain (Pfam accession number PF03734), from residues 55 to 169 (Figure 2). The NMR structure and the previously published X-ray crystal structure (Protein Data Bank [PDB] accession number 1Y7M) (Bielnicki et al., 2006) of Ldt_{Bs} are similar with rmsds of 0.85 and 0.83 Å for all heavy atoms in the LysM and catalytic domains, respectively. The relative orientation of the two domains differed in the NMR and X-ray structures by a rotation of approximately 10°.

In the LysM domain, significant differences between the NMR and X-ray structures were only detected within helices α2 and α3 (residues 29–37). The distortion of helix α2 in the X-ray structure and subsequent sliding of the backbone was attributed to crystal packing due to close intermolecular contacts involving Thr26 and Gln21. In the catalytic domain, the loops located between strands β7 and β8 (residues 107–111) and between strands β8 and β9 (residues 120–124) showed modest structural backbone reorganization in solution. While the former differences were attributed to crystal packing, the latter were related to the K120A and Q121A substitutions designed for Ldt_{Bs} crystallization (Bielnicki et al., 2006). More significant differences between the NMR and X-ray structures were localized between strands β6 and β7 (residues 88–94) and in the loop preceding catalytic Cys142 (residues 137–142). These differences were attributed to the presence of a cadmium cation and a sulfate anion in the catalytic cavity of the crystal form of Ldt_{Bs}. The Cd^{2+} ion was coordinated to the sulfate anion, to Cys142 and His126 of one subunit of a noncrystallographic dimer, and to His122 from the opposite subunit. These interactions were responsible for protein dimerization since size exclusion chromatography revealed a monomer in solution (Bielnicki et al., 2006). The monomeric state of Ldt_{Bs} was confirmed by ^{15}N relaxation measurements. In fact the correlation time τ_c = 10.4 ns calculated with TENSOR2 (Dosset et al., 2000) from relaxation data (Table S1) is consistent with the 10.6 ns predicted value computed with HYDRONMR (García de la Torre et al., 2000) based on the monomeric NMR structure. The implications of these conformational changes on the catalytic mechanism of the enzyme are discussed below.

Ldt_{Bs} Active Site Contains a pH-Independent Cys-His-Gly Catalytic Triad

A metallic ion was present in the active sites of the crystal forms of the two L,D-transpeptidases that have been analyzed by X-ray crystallography, namely Cd^{2+} in the *B. subtilis* enzyme (Ldt_{Bs}) (Bielnicki et al., 2006) and Zn^{2+} in the *E. faecium* enzyme (Ldt_{fm}) (Biarrotte-Sorin et al., 2006) (PDB accession number 1ZAT). Since the sulfur atom of the catalytic cysteine participates in the first coordination sphere of the metallic ions in Ldt_{Bs} and Ldt_{fm}, it has been proposed that the thiol groups in the crystal forms of the enzymes do not occupy a position relevant to catalysis (Biarrotte-Sorin et al., 2006; Bielnicki et al., 2006). Thus, it has been anticipated that the thiol groups could rotate in the absence of the metallic ion and point toward the imidazole ring of an invariant histidine (His126 and His421 in Ldt_{Bs} and Ldt_{fm}, respectively). Since this reorientation would place the thiol group of the catalytic Cys within hydrogen bonding distance of the imidazole ring of the conserved histidines, it was concluded that these invariant His and Cys residues may participate in catalytic triads with the main-chain carbonyl of Gly127 in Ldt_{Bs} and of Asp422 in Ldt_{fm}. According to this hypothesis, the latter group would orient the His imidazole ring in a position where it could readily capture the sulfhydryl proton, thereby generating the nucleophilic cysteinate for catalysis.

Divalent cations were absent from the buffer used in the present Ldt_{Bs} NMR study and neither the addition of EDTA or

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

Table 1. NMR and Refinement Statistics for the Structures of Ldt_{Bs} and the Acylenzyme

	Ldt_{Bs}	Ldt_{Bs} - imipenem
NMR distance and dihedral constraints		
Distance constraints		
Total NOE	3,191	2,532
Intraresidue	1,479	1,198
Interresidue	1,712	1,334
Sequential (|i – j| = 1)	681	536
Medium-range (|i – j| < 4)	325	262
Long-range (|i – j| > 5)	706	536
Total dihedral angle restraints	286	256
ϕ	143	128
ψ	143	128
Total RDCs	169	123
NH	85	79
CaHa	84	44
Q (%)[a]	17	15
Structure statistics		
Violations (mean and SD)		
Distance constraints (Å)	0.062 ± 0.005	0.060 ± 0.004
Dihedral angle constraints (°)	1.87 ± 0.03	2.15 ± 0.12
Max. dihedral angle violation (°)	17	11
Max. distance constraint violation (Å)	1.61	1.63
Deviations from idealized geometry		
Bond lengths (Å)	0.0068	0.0065
Bond angles (°)	0.97	1.34
Impropers (°)	2.34	2.76
Average pairwise rmsd (Å)[b]		
Heavy	0.70 ± 0.10 (0.57 ± 0.05)[c]	1.52 ± 0.12 (0.61 ± 0.08)[c]
Backbone	0.39 ± 0.09 (0.15 ± 0.03)[c]	0.93 ± 0.14 (0.29 ± 0.09)[c]

[a]Q is a qualitative agreement between the calculated and measured RDCs. It is evaluated as follows (Lipsitz and Tjandra, 2004):

$$Q = \sqrt{\frac{\sum_{i=1}^{n} \left(D_i^{meas} - D_i^{calc}\right)^2}{\sum_{i=1}^{n} D_i^{meas\ 2}}}.$$

[b]Pairwise rmsd was calculated among 20 refined structures in water.
[c]These values refer exclusively to rmsd calculated on secondary structure elements.

zinc chloride altered the protein [^{1}H,^{15}N-HSQC] spectrum, offering an opportunity to observe the proposed reorientation of the active site residues. In order to gain insight into the mechanism of activation of Cys142, the protonation states of histidines were determined prior to structure calculation. Assignment and chemical shift analyses of ^{1}H, ^{13}C, and ^{15}N nuclei of the four histidine residues of Ldt_{Bs} were performed with specific emphasis on the Nδ1 and Nε2 aromatic nuclei (Figure 3; Day et al., 2003; Pelton et al., 1993). These experiments revealed that His60, His122, and His146 existed at pH 6.5 as a mixture of three species in fast exchange at the NMR timescale (Table S2). On the contrary, His126 existed only as the Nδ1-H neutral tautomer, which is uncommon except for histidine residues participating in intramolecular (Pelton et al., 1993) or intermolecular (Li and Hong, 2011) hydrogen-bond networks.

In order to determine the pKa of the four Ldt_{Bs} histidines, chemical shifts of ^{1}H, ^{13}C, and ^{15}N nuclei in the imidazole rings and at the β-position were collected as a function of pH (Figures 4A and 4B). The chemical-shift pH dependence of nine nuclei per histidine was fitted simultaneously with a single pKa value (Figure S2 and Table S2). Strikingly, resonances of His126 showed no significant shift over pH values ranging from 3.6 to 10.0, indicating that neutral Nδ1-H was the predominant tautomer of His126 over the entire pH range. This observation implies a pKa lower than 3.1 for His126, which stands 3.5 units below the average value for histidine residues in the literature (Grimsley et al., 2009). Only scarce examples with extremely low pKa values for histidines have been reported in diamagnetic proteins (Yu and Fesik, 1994; Plesniak et al., 1996). Chemical shifts over the 3.6–10 pH range were also collected for the Cβ of Cys142 (Figures 4A and 4B). The chemical shifts were constant and consistent (26.0 ppm) with a sulfhydryl group (Chivers et al., 1997) over the entire pH range. These results indicate that the protonation state of Cys142 remains stable and imply a pKa outside of the explored range (below 3.1 or above 10.5). The pKa of Cys142 was first considered to be larger than 10.5. Chemical topology of histidines and Cys142 was introduced in the structure calculation in the absence of any explicit hydrogen-bond constraint. Statistics over the ten lowest-energy NMR structures establish distances of 2.57 ± 0.02 Å between Gly127-O and His126-Nδ1 and of 3.23 ± 0.05 Å between His126-Nε2 and Cys142-Sγ with a nearly optimal Sγ-Hγ-Nε2 angle of 166.9 ± 4.9°. These results precisely define the topology of a hydrogen-bond network centered on His126 (Figure 4C). Its hydrogen-bond donor capacity toward Gly127 is comforted experimentally by the detection of His126 Hδ1, which is protected from water exchange. It is balanced by a hydrogen-bond acceptor capacity at the ε2 site, which yields to a decreased pKa for the buried His126. In this geometry, the hydrogen-bond donor capacity of Cys142 is consistent with a pKa above the reported average 6.8 ± 2.7 value (Grimsley et al., 2009). The alternative hypothesis for Cys142 pKa (<3.1) failed to offer a consistent explanation for the unusual low pKa of His126. Thus, experimental data and structure calculation converge to define a catalytic triad and hydrogen-bond network involving a thiol-imidazole pair, which is predicted to be highly stable over a wide range of pH (3.6–10.0).

Noncovalent Interaction of Imipenem with a Derivative of Ldt_{Bs}

Interaction of Ldt_{Bs} with β-lactams is expected to involve three steps including formation of a noncovalent complex between the enzyme and the drug, formation of a thioester bond with the catalytic cysteine, and, finally, hydrolysis of the resulting acylenzyme to form native Ldt_{Bs} and inactive drug (Figure 5). In order to investigate the first step of the reaction, the catalytic Cys

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

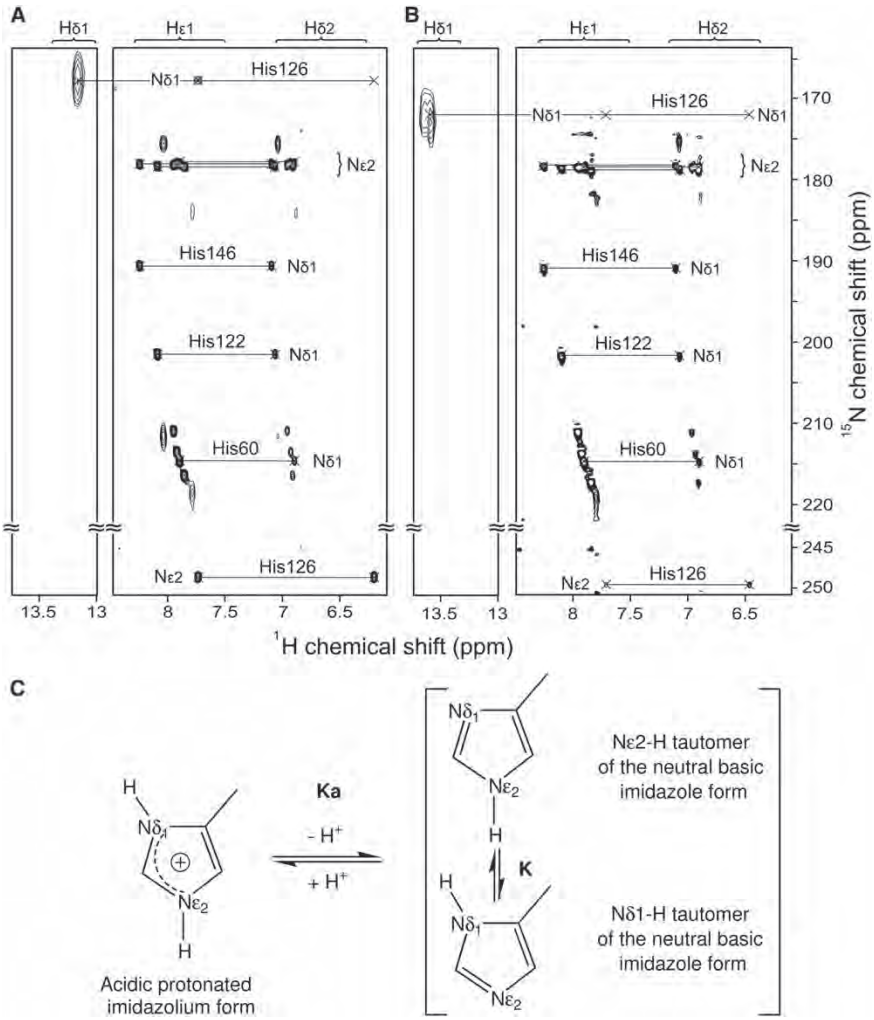


Figure 3. Determination of Histidines Protonation State at pH 6.5

[^{1}H-^{15}N]-correlation experiments optimized for the detection of $^1J_{NH}$ (left panel) and $^2J_{NH}$ /$^3J_{NH}$ (right panel) couplings in histidine side chains recorded on (A) Ldt_{Bs} and (B) Ldt_{Bs} acylated by imipenem at 25°C (pH 6.5). Unassigned peaks correspond to the hexahistidine tag. Only the imidazole ring of His126 exists as the Nδ1-H tautomer, suggesting involvement in hydrogen bonding. (C) Acido-basic and tautomeric equilibria for the histidine imidazole ring.
See also Table S2.

residue was replaced by Ala to prevent formation of the acylenzyme. The C142A substitution did not alter the chemical shifts of the backbone resonances observed in the [^{1}H,^{15}N]-HSQC spectrum (Figure S1). Since minor effects were only seen in residues surrounding position 142, the substitution did not modify the conformation of the protein.

NMR chemical shift perturbation experiments were performed by incubating Ldt_{Bs} C142A with increasing concentrations of imipenem to observe the formation of the noncovalent complex (E.I, Figure 5A). A fast exchange regime was detected between the apoenzyme and the complex. The [^{1}H,^{15}N]-chemical shift perturbations increased with the concentration of imipenem up to the highest drug concentration that could be tested (37.5 mM) (Figure S3). The absence of saturation of Ldt_{Bs} by imipenem indicates that Ldt_{Bs} C142A displays low affinity for imipenem and provides an estimate of the minimum value of the dissociation constant for the noncovalent complex (K_D > 37 mM) (Figure S4A). The residues that were affected by drug binding were mostly located at the surface of the protein in the vicinity of catalytic Cys142 (Figure 5B and Figure S3). Thus, imipenem specifically bound with a low affinity to the catalytic pocket of Ldt_{Bs} C142A in the absence of enzyme acylation.

Kinetic analyses of wild-type Ldt_{Bs} acylation by stopped-flow spectrophotometry (Figure S4B) revealed that the pseudo first order rate constant of the acylation reaction was proportional to the concentration of imipenem in the range of drug concentration (0–400 μM) that was experimentally accessible using this technique. Thus, saturation of the wild-type enzyme was not

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

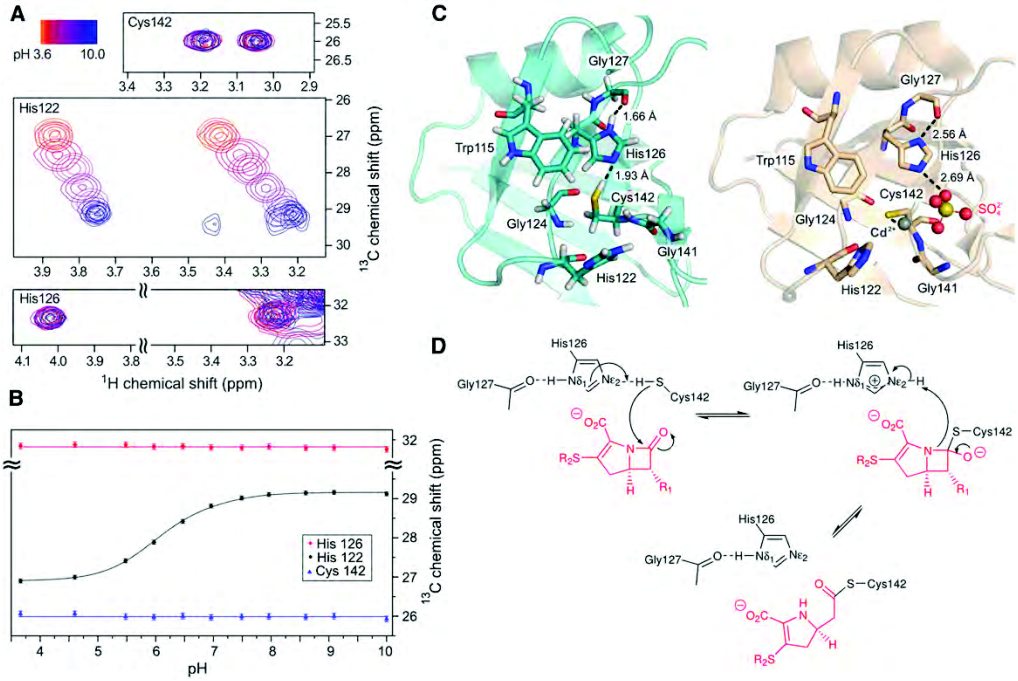


Figure 4. Analyses of the Topology of the Ldt_{Bs} Catalytic Triad

(A) Overlay of portions of 2D [^{1}H,^{13}C]-HSQC spectra recorded over the 3.6–10.0 pH range for the ^{1}Hβ-^{13}Cβ correlations for Cys142, His122, and His126.

(B) pH dependence of ^{13}Cβ NMR chemical shifts showing that the protonation state of Cys142 (in blue) and His126 (in red) remains the same over the 3.6–10.0 pH range. His122 (in black) is titrated.

(C) Topology of active site residues and depiction of the hydrogen-bond network within the catalytic triad in Ldt_{Bs} NMR structure (blue teal, left). The topology of active sites residues of the X-ray structure (wheat) is included on the right to highlight reorientation of the catalytic Cys142 in the presence of Cd^{2+}.

(D) Deduced role of the catalytic triad in the acylation of Ldt_{Bs} by carbapenems.

See also Figure S2.

observed in kinetic analyses indicating that the low affinity of Ldt_{Bs} C142A for imipenem deduced from NMR experiments (above) was not merely the consequence of the substitution introduced into the active site to block acylation of Ldt_{Bs} by the drug. Although kinetic and NMR analyses explore different ranges of ligand concentration (up to 0.4 and 37.5 mM, respectively) and provide different evaluations of the binding step (K_{app} and K_D, respectively), both approaches suggest that the affinity of Ldt_{Bs} for imipenem is at best in the millimolar range.

Since Ldt_{Bs} does not form covalent adducts with ampicillin, formation of a noncovalent complex involving the wild-type enzyme was tested with this β-lactam. Ldt_{Bs} C142A was included in this analysis for comparison (Figures S3 and S4A). NMR chemical shift perturbation experiments revealed a fast exchange regime between the apoenzymes and the complexes. The residues affected by ampicillin in Ldt_{Bs} C142A and Ldt_{Bs} were similar to those affected by imipenem in Ldt_{Bs} C142A. Binding of ampicillin to both forms of the enzyme occurred with similar low affinities ($K_D \geq 100$ mM), which could not be precisely determined since they were in the order of the solubility of the drug. Together, these results indicate that imipenem and ampicillin bind with similar low affinities to the L,D-transpeptidase.

Acylation by Imipenem Induces a Structural Rearrangement of Ldt_{Bs}

Ldt_{Bs} was incubated with imipenem in a one-to-one molar ratio and multidimensional NMR spectroscopy was performed in the conditions previously used for the apoenzyme. The acylenzyme formed within a minute and remained stable for at least five days. Comparison of [^{1}H,^{15}N]-HSQC spectra of the two forms of the enzyme showed significant differences that revealed structural rearrangement of the enzyme upon acylation (Figure S5). The neutral Nδ1-H tautomer remained the only acido-basic form of catalytic His126 at pH 6.5 (Figure 3B). While the pKa of this residue could not be determined in the acylenzyme, the neutral Nδ1-H tautomer appears less stable than in the apoenzyme, as none of the side-chain correlations could be detected at pH 5.5. The structure of the acylenzyme at pH 6.5 with an explicitly modified imipenem-bound cysteine residue (Figure 6) was calculated using 2,532 unambiguous distance restraints derived from NOE, 128 dihedral constraints derived from chemical-shift data, and 123 angular restraints derived from residual dipolar couplings (Table 1). The rmsd calculated for the 20 lowest energy structures with respect to the average structure was 0.61 Å for heavy atoms of residues participating in α helices and β sheets. These results indicate that the secondary structure elements of

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

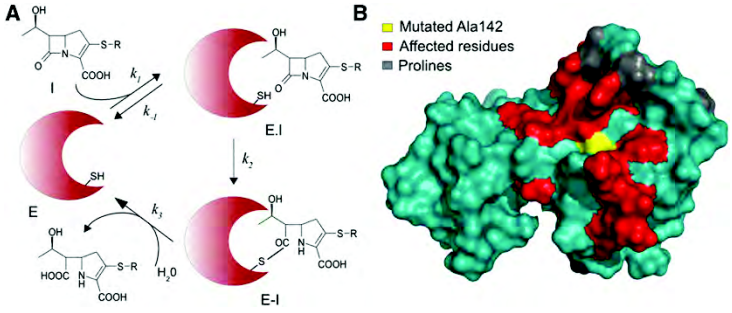


Figure 5. Interaction of Imipenem with Ldt_{Bs} and Ldt_{Bs} C142A Deficient in Acylenzyme Formation

(A) Schematic representation of a complete catalytic cycle. The first step corresponds to binding of imipenem (I) to enzyme (E) leading to the formation of a noncovalent complex (E.I). The derivative of Ldt_{Bs} with the C142A substitution is blocked at this step. The second step corresponds to the chemical step of the reaction leading to acylation of Cys142 by imipenem. This reaction involves the nucleophilic attack of the β-lactam carbonyl by the sulfur of Cys142. The resulting acylenzyme (E.I) contains imipenem with an opened β-lactam ring covalently linked to Cys142 through a thioester bond. In the third step, the acylenzyme is eventually hydrolyzed to produce apoenzyme and inactivated imipenem. This reaction product was only detected after several days.

(B) Surface representation of chemical shift perturbations showing residues affected by noncovalent binding of imipenem to Ldt_{Bs} C142A. Residues with chemical shift differences greater than 0.0125 ppm are shown in red. The active-site cysteine is shown in yellow.

See also Figures S3 and S4.

the acylenzyme were well defined. Comparison of [^{1}H,^{15}N]-HSQC spectra of the apo- and the acylenzyme revealed that the intensity of some of the resonances in the catalytic domain of Ldt_{Bs} were decreased, down to extinction in some cases (Figures 7A and 7B). Resonances that disappeared or decreased in intensity were mostly located in the vicinity of the catalytic cysteine in three regions of the protein, which comprised residues 84–97 (region 1), 119–126 (region 2), and 136–148 (region 3). A unique structure could not be defined for these regions, which account for the significant increase in the rmsd value for the acylenzyme (1.52 Å, Figures 6B and 6D) relative to the apoenzyme (0.70 Å, Figures 6A and 6C). In the 20 lowest-energy calculated structures, the bound imipenem ligand samples a wide range of orientations, consistent with its exposure to the solvent and local dynamics.

Flexibility Is Induced by Acylation of Ldt_{Bs} by Imipenem

^{15}N relaxation NMR experiments (Farrow et al., 1994; Mittermaier and Kay, 2009) were performed to compare the dynamical properties of the apo- and acylenzymes. The backbone of the apo form of Ldt_{Bs} was rigid, with an average heteronuclear Overhauser effect (hetNOE) value of 0.80 ± 0.02 for all regions except the termini (residues 1–5 and 164–169) and two loops (residues 35–41 and 107–110). Both forms of Ldt_{Bs} gave similar R_1 and R_2 relaxation rate constants and overall hetNOE values (Table S1). The main difference between the apo form and the acylenzyme was a decrease in the intensity of resonances in regions 1–3, as described above. This decrease was likely to result from increased conformational exchange in the millisecond timescale. To further investigate this conformational exchange and to gain access to the timescales of the motion, relaxation-dispersion experiments were performed to evidence dynamics at the micro- to millisecond timescales. Transverse relaxation rate constants (R_2) were measured for different repetition numbers of Carr-Purcell-Meiboom-Gill (CPMG) pulse schemes (Wang et al., 2001). This approach did not reveal any conformational exchange in the apoenzyme confirming the rigidity of this form. In the acylenzyme, conformational exchange was detected for residues 84, 85, 121, 126, 141, and 146 (Figure 7 and Figure S6). Simple two-site exchange fitting procedures (see Experimental Procedures) gave exchange rate constants of 2.0 ± 1.0, 2.3 ± 0.2, and 2.2 ± 0.2 ms^{-1} for residues in, or in the immediate proximity of, region 1 (Ile84 and Ala85), region 2 (His126), and region 3 (Gly141), respectively. Thus, acylation of Ldt_{Bs} induced conformational dynamics in the millisecond regime in the vicinity of the catalytic cysteine.

Exploring Motional Timescales for Different Covalent Adducts

In order to investigate the basis for increased dynamics upon acylation of Ldt_{Bs} by imipenem, 2-nitro-5-thiobenzoate (TNB) was covalently linked to the enzyme through a disulfide bond, following reaction of the lone catalytic cysteine with 5-5′-dithio-bis-(2-nitrobenzoic acid) (DTNB). The rationale for this study was to compare the impact of covalent modifications on the catalytic cysteine by two structurally unrelated groups, imipenem (through a thioester linkage) and TNB (through a disulfide linkage). For this purpose, the ^{13}C,^{15}N-Ldt_{Bs} disulfide derivative was prepared and a [^{1}H,^{15}N]-HSQC spectrum was recorded and compared to those of the apoenzyme and of the acylenzyme obtained with imipenem (Figure S5A). After backbone amide resonance assignments through conventional 3D experiments, a set of residues was identified that showed significant signal intensity decrease (Figure 7A). Strikingly, this set of residues was almost identical to that affected by the acylation of Ldt_{Bs} by imipenem. However, comparison of the two adducts revealed differences in the type of modifications in the [^{1}H,^{15}N]-HSQC spectra. Whereas several residues were undetectable in the imipenem adduct, only a minority of these residues were missing in the disulfide adduct, although a significant signal intensity decrease was detected for most of them. Flexibility of these residues was investigated by recording ^{15}N relaxation NMR experiments on the TNB adduct. Similar overall relaxation parameters were observed for the three forms of the enzyme except for regions 1, 2, and 3 (Table S1). In comparison to the apoenzyme, formation of the disulfide bridge induced increase in R_2 relaxation rate constants ($R_{2,ex}$) for residues 86, 88, 91, 93, 95, and 96 in region 1, 121, and 123 in region 2, and 137, 138, 141, 143, 144, 145, and 146 in region 3. Relaxation-dispersion experiments recorded for the TNB adduct revealed conformational exchange in the micro- to millisecond regime for those residues (Figure 7C and Figure S6). Two-site exchange fitting

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

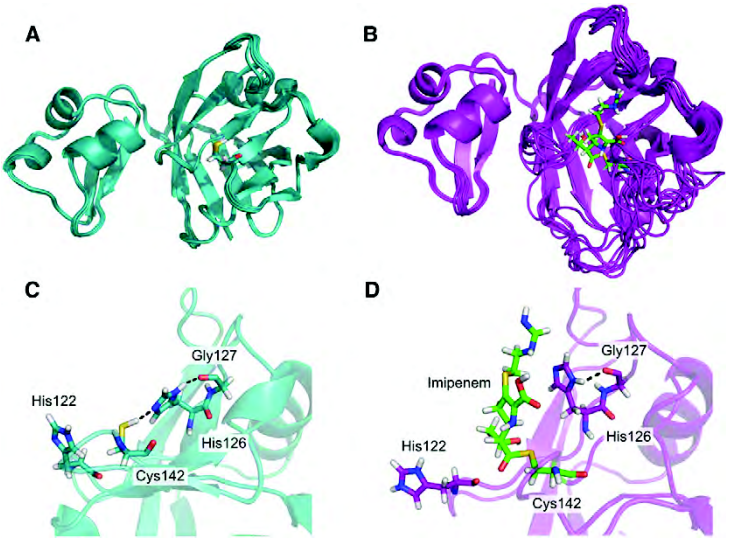


Figure 6. Structural Rearrangement Resulting from the Acylation of Ldt_{Bs} by Imipenem
Superposition of the 10 lowest energy structures is shown for Ldt_{Bs} (A) and for the acylenzyme (B). For the acylenzyme, the orientation of Cys142 bound to imipenem (in green sticks) is only depicted for the lowest-energy structure and does not take into account the unrestricted conformational sampling of this modified residue. Lower panels allow comparison of the lowest energy structure of apoenzyme (C) and acylenzyme (D) active sites.
See also Figure S5.

procedures of $R_{2,ex}$ gave exchange rate constants (k_{ex}) of 6.1 ± 1.4, 7.0 ± 3.0, and 6.3 ± 1.8 ms^{-1} for regions 1 (Val86, Lys88, Thr91, and Thr93), 2 (Tyr123), and 3 (Val138, Gly141, Arg144, and Met145), respectively. Together these results show that covalent modification of the active site cysteine of Ldt_{Bs} by formation of a disulfide bond with TNB or by acylation with imipenem-induced conformational flexibility in the same regions of the protein. The exchange rates were approximately three times faster for the TNB adduct than for the acylenzyme. Imipenem induced larger conformational exchange effects than TNB as shown for representative residues in Figure 7C. As outlined for residue Phe111 in this figure, exchange was not detected outside from the flexible regions for any of the three forms of the enzyme. For residues located near the border of the flexible regions, exchange was detected in the acylenzyme but not in the apo form or in the TNB adduct (e.g., residue His126). In the core of the flexible regions, certain residues displayed a higher exchange rate in the acylenzyme than in the TNB adduct (Gly141). Other residues showed a high relaxation rate in the TNB adduct, whereas a greater relaxation rate may account for the absence of resonance in the acylenzyme (Thr91). Thus, modifications of Ldt_{Bs} induced by imipenem and TNB involved an increase in the dynamics of the protein at a millisecond timescale that affected the same regions located in the vicinity of the catalytic residue.

DISCUSSION

Structural data on L,D-transpeptidases are limited to two reports of crystal structures despite the relevance of this unique family of peptidoglycan polymerases as targets for the development of new drugs active in the treatment of multidrug resistant tuberculosis (Gupta et al., 2010; Hugonnet et al., 2009; Lavollay et al., 2008). The X-ray structure (Figures 2 and 4C) of the *B. subtilis* enzyme, Ldt_{Bs}, has generated limited understanding of the active site since the conformations of the catalytic Cys142 and the loop connecting strands β6 and β7 were altered by the artifactual presence of cadmium and sulfate ions (Bielnicki et al., 2006). Here, we solved the NMR structure of the apo form of the enzyme (Figure 2) and showed that residues Cys142, His126, and Gly127 form a catalytic triad (Figure 4C). Acylation of Ldt_{Bs} by imipenem induced important structural and dynamic rearrangements that may account for previous unsuccessful attempts to solve the crystal structure of covalent adducts.

The hydrogen-bond network of the triad remained stable from pH 3.6 to 10.0 implying pKa values below 3.1 and above 10.5 for His126 and Cys142, respectively. Since the enzyme fraction presenting an imidazolium-thiolate ion pair is estimated to be less than 0.02% ($10^{-\Delta pKa/2}$), a proton shuttle mechanism similar to that of classic Ser-His-Asp serine proteases is proposed (Figure 4D; Harris and Turner, 2002). According to this mechanism, His126 captures the Sγ proton of Cys142 to generate the nucleophilic thiolate that attacks the carbonyl carbon of the β-lactam ring. In the next step, His126 may release its acidic proton to the nitrogen of the β-lactam ring in agreement with the exclusive presence of the neutral Nδ1-H tautomer in the acylenzyme.

Ldt_{Bs} differs from most cysteine peptidases (Harris and Turner, 2002) by the absence of an imidazolium-thiolate ion pair in the apo form of the enzyme. The lack of a preexisting thiolate may not imply a profound difference in the catalytic mechanism since recent studies in the papaïn-like protein family C1 and sortases suggest that a neutral imidazole-thiol pair forms upon ligand binding (Connolly et al., 2003; Schneck et al., 2008; Shokhen et al., 2011). Since peptidoglycan polymerization occurs in the periplasm, the cytoplasmic membrane does not protect L,D-transpeptidases from environmental oxidizing agents. Thus, the absence of Cys142 activation as a thiolate in the absence of substrate may be physiologically significant for this family of peptidoglycan polymerases.

NMR has been successfully applied to molecular systems with significant intrinsic mobility, and NMR relaxation tools have been developed in order to explore conformational dynamics over a wide range of timescales (Mittermaier and Kay, 2009). These methods have profoundly changed our view on the role of flexibility in molecular recognition processes (Baldwin and Kay, 2009). In the case of Ldt_{Bs}, several models need to be considered in order to understand the large structural rearrangement and conformational exchange processes induced by acylation. The lock-and-key model can be excluded since it does not

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

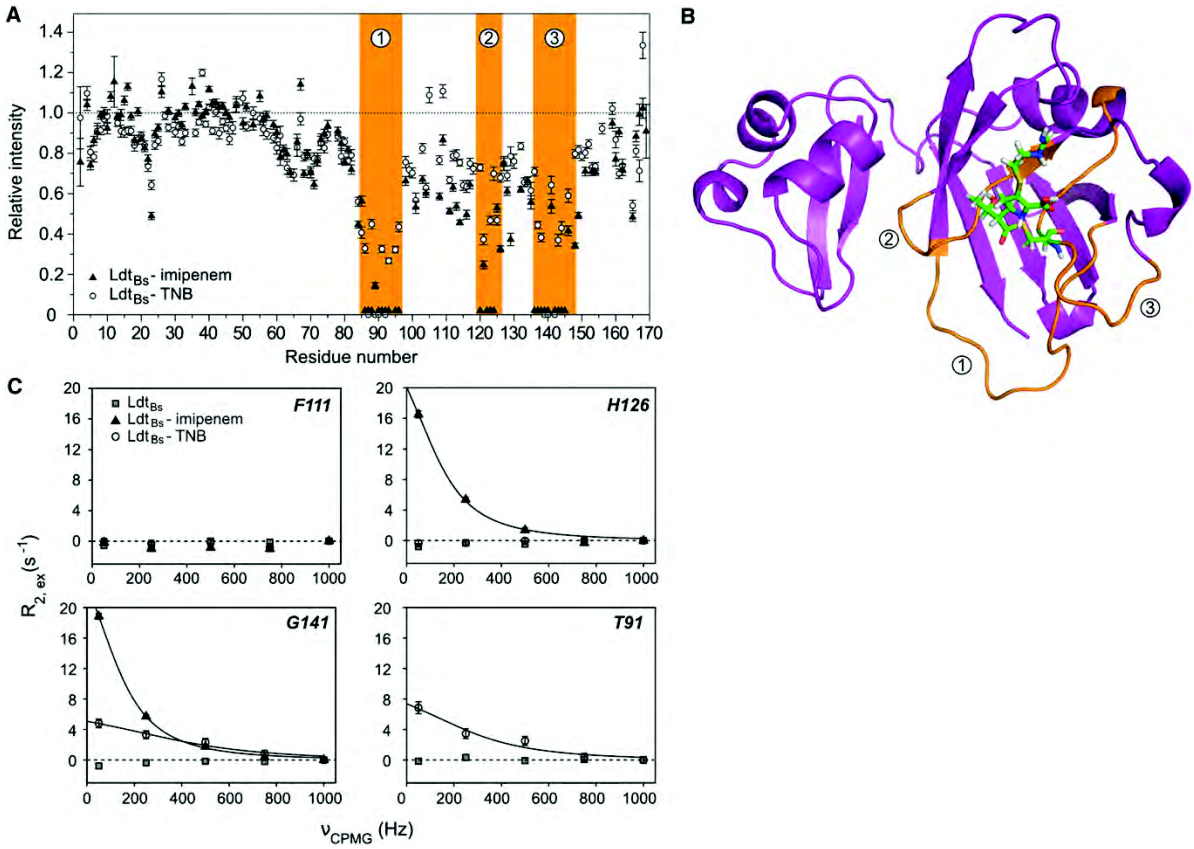


Figure 7. Impact of Covalent Binding of Imipenem and TNB on Ldt$_{Bs}$ Flexibility

(A) Intensities in [^{1}H,^{15}N]-HSQCs of Ldt$_{Bs}$ covalent adducts with imipenem (▲) and TNB (O) relative to the native enzyme.

(B) The three regions mostly affected by imipenem and TNB binding are colored in orange on the structure of Ldt$_{Bs}$ acylenzyme and are located in the vicinity of the imipenem-bound Cys142 (shown as green sticks in the conformation of the lowest energy structure).

(C) Relaxation-dispersion $R_{2,ex}$ contributions as a function of CPMG repetition frequencies (ν_{CPMG}) for representative residues of the three regions (Thr91 for region 1, His126 for region 2, and Gly141 for region 3) and for a residue external to these regions (Phe111). For each of these residues, experimental data points are shown for native Ldt$_{Bs}$ (■), for the acylated enzyme (▲), and for the TNB adduct (O). Solid lines represent fits with a simple two-site model of $R_{2,ex}$ as a function of ν_{CPMG}. Dashed lines represent the expected behavior in the absence of conformational exchange. In the apo form of Ldt$_{Bs}$, none of the residues showed exchange contribution. In the acylenzyme, exchange is only detected in regions 1 to 3 and strong exchange contributions are likely to account for missing resonances (e.g., Thr91). $R_{2,ex}$ contributions for the TNB adduct occur in the same regions, but with smaller magnitude.

In (A) and (C), error bars represent absolute propagated uncertainties calculated using partial derivatives and peak height errors estimated using NMRPipe peak analysis routines (Delaglio et al., 1995).

See also Figure S6 and Table S1.

take into account conformational changes. The "induced fit" model (Koshland, 1958) proposes that drug binding drives the protein toward a new state. In the "selected fit" model, a weakly populated ligand-free conformation is selected upon drug binding with a subsequent population shift toward the corresponding conformers (Frauenfelder et al., 1991; Hammes et al., 2009; Tsai et al., 1999). As discussed below, neither model fully accounts for the structure and dynamics of the different forms of Ldt$_{Bs}$.

Several lines of evidence indicate that the "selected fit" model does not apply to the interaction of Ldt$_{Bs}$ with β-lactams. First, the relaxation dispersion data recorded on the apoenzyme failed to reveal any preexisting conformational exchange in the millisecond timescale. Second, acylation of Ldt$_{Bs}$ by imipenem did not result in stabilization of a weakly populated conformation that might have remained undetected. Instead, acylation of Ldt$_{Bs}$ by imipenem induced backbone flexibility in the vicinity of the binding site, as established by relaxation-dispersion measurements (Figure 7). Third, noncovalent binding of imipenem to Ldt$_{Bs}$ C142A did not reveal any conformational exchange. Instead, NMR chemical-shift perturbations indicated a fast exchange regime, and residues affected by drug binding were located at the surface of the protein, close to position 142. In contrast to Ldt$_{Bs}$, a "selected fit" model is supported by structural analyses of classical D,D-transpeptidases, as, for example, in PBP1b from *Streptococcus pneumoniae*, which

Structure
β-Lactams Induce Dynamics in L,D-Transpeptidases

has been crystallized in the apoenzyme form and as covalent adducts resulting from acylation of the catalytic serine by β-lactams (Lovering et al., 2006; Macheboeuf et al., 2005) and peptidoglycan precursors (Macheboeuf et al., 2008). Since these structures suggest that "closed" and "open" conformations of *S. pneumoniae* PBP1b could coexist in solution, a "selected fit" model appears to apply to the conformational modification observed upon acylation and could be the basis for the proposed activation of PBP1b upon recognition of the substrate. Thus, Ldt_{Bs} appears to be radically different from PBP1b, both because the catalytic cysteine is easily accessible and, as stated above, β-lactam binding is not associated with stabilization of any preexisting conformation.

Since the selected fit model does not apply to Ldt_{Bs}, conformational variations in the three forms of the enzymes should be considered in the framework of the induced fit model, as, for example, in PBP2a from methicillin-resistant *Staphylococcus aureus* (Lim and Strynadka, 2002). Since the acylenzyme obtained with imipenem existed as multiple conformers in rapid exchange (rate constant ~2 ms^{-1}), the energy landscape for this form of the protein appeared to be flat. Flexibility is also observed for the catalytic cysteine, indicating the absence of any preferential orientation of the drug and of any enthalpically stabilized drug-protein interactions. Covalent binding of TNB to Cys142 induced qualitatively similar perturbations in the same three regions of the protein (Figure 7) despite the lack of structural similarity between TNB and imipenem. Thus, the role of the free sulfhydryl group in stabilizing the conformation of the native enzyme is worth examining. It could, for example, be argued that the Sγ-Hγ–Nε2 hydrogen interaction between Cys142 and His126 (Figure 4C) is essential to maintain the conformation of the apo form of the enzyme. This would account for the fact that disruption of the hydrogen interaction by formation of thioester or disulfide bonds with dissimilar ligands have similar effects. This explanation was ruled out since the C142A substitution had essentially no impact on the conformation of the apoenzyme. Alternatively, steric hindrance could be the driving force for increased protein flexibility since this would account for the different behaviors of the Ldt_{Bs} C142A mutant and the adducts. However, as stated above, the catalytic Cys142 is freely accessible in the acylenzyme, and the imipenem-bound cysteine can be modeled at the surface in many different orientations without steric conflict (Figure 6D).

L,D-transpeptidases are highly specific for carbapenems as these enzymes are rapidly acylated by this class of β-lactams (Mainardi et al., 2007; Figure 1D) but not by penicillins, such as ampicillin. To investigate the basis for this specificity, NMR chemical-shift perturbations were performed for noncovalent adducts containing Ldt_{Bs} and ampicillin, or Ldt_{Bs} C142A and imipenem (Figure S3). In both cases, the complex and the apoenzyme were in rapid equilibrium, the same residues were affected by drug binding, and increased flexibility was not observed upon ligand binding. These results confirm that increased flexibility requires covalent modification of Cys142 since conformational changes are not induced by the presence of β-lactams in the active site in the absence of acylation. Since ampicillin and imipenem interacted with Ldt_{Bs} with similar low affinities (Figure S4A), the specificity of the enzyme for carbapenems is merely determined by the chemical step of the reaction, probably at the activation step of Cys142 by the drug. Analyses of the chemical step of the reaction by quantum chemistry will therefore be required to understand the basis of L,D-transpeptidase specificity for carbapenems.

EXPERIMENTAL PROCEDURES

Plasmids and Proteins

Details of cloning and of protein production, modification, and purification are described in Supplemental Experimental Procedures.

NMR Spectroscopy

Unless otherwise noted, NMR samples were prepared in 12.5 mM MES and 150 mM NaCl buffer at pH 6.5 with 10% D_2O. Data for structure determination were collected on 0.7 mM ^{13}C,^{15}N-labeled protein samples using conventional approaches (Sattler et al., 1999) on Bruker and Agilent spectrometers operating at 600, 800, and 900 MHz proton frequencies, as described in Supplemental Experimental Procedures. NMR data were processed with NMRPipe (Delaglio et al., 1995) and analyzed using the CcpNmr Analysis routines of CCPN (Vranken et al., 2005).

Structural Restraints

Secondary structure elements were identified by analysis of ^{13}C secondary chemical shifts and phi/psi dihedral angular restraints were derived with TALOS+ (Shen et al., 2009). Distance restraints from 3D ^{13}C-methyl-selective-NOESY experiments (Van Melckebeke et al., 2004) and from a 2D homonuclear [^1H,^1H]-NOESY were manually assigned and extracted from peak volumes measured in CcpNmr. The automatic peak-picking and NOE assignment of the ^{13}C- and ^{15}N-NOESY-HSQC spectra were performed using the iterative procedure of UNIO'10 version 2.0.1, based on the CYANA/ATNOS/CANDID suite of programs (Herrmann et al., 2002). RDCs were analyzed as detailed in Supplemental Experimental Procedures.

Structure Calculation and Refinement

ARIA 2.3 (Rieping et al., 2007) was run with eight iterations including 100 structures per iteration, with the exception of the last cycle run with 1,000 structures. The 20 structures with lowest energy were subsequently refined with crystallography and NMR system (CNS) (Brünger et al., 1998) using explicit solvent in a molecular dynamics simulation. CNS and PROCHECK-NMR were used for analysis of the structural ensembles and quality assessment of the structural data. For the acylenzyme, the covalent binding of imipenem was explicitly introduced in the structure calculation using a modified cys-residue created with PRODRG (Schüttelkopf and van Aalten, 2004). The Ramachandran plot showed 83.5% of the residues in most favored regions, 15.3% and 1.2% in additional allowed and generously allowed regions, and 0.0% in disallowed regions for Ldt_{Bs}. For the acylenzyme, residues were located at 78.1% in most favored regions, 18.3% and 2.0% in additional allowed and generously allowed regions, and 1.6% in disallowed regions. Structures were visualized and figures prepared with PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.

NMR Titrations

Increasing molar ratios of imipenem and ampicillin (up to 300 and 1,000 equivalents, respectively) were added to ~150 μM solutions of ^{15}N-labeled Ldt_{Bs} or Ldt_{Bs} C142A at pH 6.5 and 298 K. Chemical shift perturbations were monitored through the comparison of 2D [^1H,^{15}N]-HSQC spectra recorded at 600 MHz proton frequency and analyzed to extract structural and thermodynamics binding information as detailed in Supplemental Experimental Procedures.

^{15}N-Relaxation

Residue-specific backbone amide ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates and steady-state heteronuclear {^1H}-^{15}N-NOE were collected at 600 MHz proton frequency and 298 K on the apo form, the acylenzyme and the TNB adduct of Ldt_{Bs}. The CPMG repetition rate used to collect ^{15}N R_2 relaxation data was set to 400 Hz. Resonances were automatically picked and fitted simultaneously on all spectra using the NlinLS procedure in NMRPipe (Delaglio et al., 1995). Peak intensities were then fitted using a

Structure

β-Lactams Induce Dynamics in L,D-Transpeptidases

Monte-Carlo algorithm with a monoexponential decay model. The hetNOE ratio was determined directly after fitting of the resonances and extraction of the intensities in each spectrum. Rotational correlation times (10.0, 10.4, and 11.1 ns for Ldt_{Bs}, Ldt_{Bs}-TNB, and Ldt_{Bs}-imipenem, respectively) were obtained from a Lipari-Szabo model-free analysis conducted under TENSOR2 (Dosset et al., 2000) after exclusion of residues with hetNOE values lower than 0.65, or with R_1 or R_2 values exceeding one standard deviation from the mean.

Relaxation-Dispersion

On each of the Ldt_{Bs} samples (apo form, acylenzyme and TNB adduct) a series of BEST-modified (Lescop et al., 2010) CPMG experiments (Wang et al., 2001) was recorded at 600 MHz proton frequency and 298 K with a constant relaxation delay T_{relax} = 40 ms and a number of ^{15}N-CPMG π pulses varying from 4 to 80. For each backbone amide exchange contribution on the ^{15}N transverse relaxation correlation, $R_{2,ex}$, as a function of ν_{CPMG} was calculated as:

$$R_{2,ex}(\nu_{CPMG}) = -\frac{1}{T_{relax}}\ln\left(\frac{I(\nu_{CPMG})}{I_{ref}}\right), \quad (1)$$

where I_{ref} is the peak intensity in the spectrum collected with ν_{CPMG} = 1,000 Hz and $I(\nu_{CPMG})$ is the peak intensity for a particular CPMG frequency. Residues with $R_{2,ex}$ values greater than 3 s^{-1} for low ν_{CPMG} frequencies were classified as exhibiting significant exchange and localized in three clearly identified regions of the imipenem and TNB Ldt_{Bs} adducts. For these residues the $R_{2,ex}$ dependence as a function of ν_{CPMG} was least-square fitted in the frame of a two-site exchange model (Palmer, 2004):

$$R_{2,ex}(\nu_{CPMG}) = \frac{\varphi_{ex}}{k_{ex}}\left[1 - \frac{2\tanh\left(\frac{k_{ex}}{4}\left(\frac{1}{2\nu_{CPMG}} - t_{180}\right)\right)}{\frac{k_{ex}}{2}\left(\frac{1}{2\nu_{CPMG}} - t_{180}\right)}\right], \quad (2)$$

where k_{ex} is the exchange rate constant between the two sites, φ_{ex} is a per-residue adjustable parameter that relates to the population of each site and the difference in the NMR spectral properties of the two sites, and t_{180} is the duration of the ^{15}N π pulse in the NMR experiment. As residues from the same region showed similar exchange rates, the final fits were performed with a single k_{ex} value per region.

ACCESSION NUMBERS

Coordinates of twenty structures of the apo and acylated forms of Ldt_{Bs} have been deposited in the Protein Data Bank under the accession codes 3ZQD and 4A52. Chemical shifts, residual dipolar couplings, and relaxation data have been deposited in the BioMagResBank under accession numbers 17701 and 18037.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.str.2012.03.015.

ACKNOWLEDGMENTS

We thank B. Brutscher, A. Dessen, M. Field, P. Gans, and J.-L. Mainardi for critical reading of the manuscript. We thank T. Kern for the initial assignment of an Ldt_{Bs} mutant. Financial support by the Access to Research Infrastructures activity in the 6th-Framework Programme of the EC (Contract #RII3-026145, EU-NMR) for conducting the research on the 900 MHz spectrometer of the CERM Facility and by the French TGIR-RMN for conducting the research on the 800 MHz spectrometer of the ICSN Facility are gratefully acknowledged. This work was supported by contracts PeptidoNMR and CarbaTub from the Agence Nationale de la Recherche.

Received: December 8, 2011
Revised: February 29, 2012
Accepted: March 5, 2012
Published: May 8, 2012

REFERENCES

Baldwin, A.J., and Kay, L.E. (2009). NMR spectroscopy brings invisible protein states into focus. Nat. Chem. Biol. *5*, 808–814.

Biarrotte-Sorin, S., Hugonnet, J.E., Delfosse, V., Mainardi, J.L., Gutmann, L., Arthur, M., and Mayer, C. (2006). Crystal structure of a novel beta-lactam-insensitive peptidoglycan transpeptidase. J. Mol. Biol. *359*, 533–538.

Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z.S. (2006). *B. subtilis* ykuD protein at 2.0 A resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. Proteins *62*, 144–151.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Chivers, P.T., Prehoda, K.E., Volkman, B.F., Kim, B.M., Markley, J.L., and Raines, R.T. (1997). Microscopic pKa values of *Escherichia coli* thioredoxin. Biochemistry *36*, 14985–14991.

Connolly, K.M., Smith, B.T., Pilpa, R., Ilangovan, U., Jung, M.E., and Clubb, R.T. (2003). Sortase from *Staphylococcus aureus* does not contain a thiolate-imidazolium ion pair in its active site. J. Biol. Chem. *278*, 34061–34065.

Day, R.M., Thalhauser, C.J., Sudmeier, J.L., Vincent, M.P., Torchilin, E.V., Sanford, D.G., Bachovchin, C.W., and Bachovchin, W.W. (2003). Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN' by NMR. Protein Sci. *12*, 794–810.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J. Biomol. NMR *16*, 23–28.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry *33*, 5984–6003.

Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. Science *254*, 1598–1603.

García de la Torre, J., Huertas, M.L., and Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J. Magn. Reson. *147*, 138–146.

Grimsley, G.R., Scholtz, J.M., and Pace, C.N. (2009). A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci. *18*, 247–251.

Gupta, R., Lavollay, M., Mainardi, J.L., Arthur, M., Bishai, W.R., and Lamichhane, G. (2010). The Mycobacterium tuberculosis protein LdtMt2 is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. Nat. Med. *16*, 466–469.

Harris, T.K., and Turner, G.J. (2002). Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. IUBMB Life *53*, 85–98.

Hammes, G.G., Chang, Y.C., and Oas, T.G. (2009). Conformational selection or induced fit: a flux description of reaction mechanism. Proc. Natl. Acad. Sci. USA *106*, 13737–13741.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J. Biomol. NMR *24*, 171–189.

Hugonnet, J.E., Tremblay, L.W., Boshoff, H.I., Barry, C.E., 3rd, and Blanchard, J.S. (2009). Meropenem-clavulanate is effective against extensively drug-resistant *Mycobacterium tuberculosis*. Science *323*, 1215–1218.

Koshland, D.E. (1958). Application of a theory of enzyme specificity to protein synthesis. Proc. Natl. Acad. Sci. USA *44*, 98–104.

Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., Blanot, D., Gutmann, L., and Mainardi, J.L. (2008). The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L,D-transpeptidation. J. Bacteriol. *190*, 4360–4366.

Structure

β-Lactams Induce Dynamics in L,D-Transpeptidases

Lavollay, M., Fourgeaud, M., Herrmann, J.L., Dubost, L., Marie, A., Gutmann, L., Arthur, M., and Mainardi, J.L. (2011). The peptidoglycan of *Mycobacterium abscessus* is predominantly cross-linked by L,D-transpeptidases. J. Bacteriol. *193*, 778–782.

Lescop, E., Kern, T., and Brutscher, B. (2010). Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type ^{1}H-^{15}N correlation experiments. J. Magn. Reson. *203*, 190–198.

Li, S.H., and Hong, M. (2011). Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR. J. Am. Chem. Soc. *133*, 1534–1544.

Lipsitz, R.S., and Tjandra, N. (2004). Residual dipolar couplings in NMR structure analysis. Annu. Rev. Biophys. Biomol. Struct. *33*, 387–413.

Lim, D., and Strynadka, N.C. (2002). Structural basis for the β lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. Nat. Struct. Biol. *9*, 870–876.

Lovering, A.L., De Castro, L., Lim, D., and Strynadka, N.C. (2006). Structural analysis of an "open" form of PBP1B from *Streptococcus pneumoniae*. Protein Sci. *15*, 1701–1709.

Macheboeuf, P., Di Guilmi, A.M., Job, V., Vernet, T., Dideberg, O., and Dessen, A. (2005). Active site restructuring regulates ligand recognition in class A penicillin-binding proteins. Proc. Natl. Acad. Sci. USA *102*, 577–582.

Macheboeuf, P., Lemaire, D., Teller, N., Martins, Ados.S., Luxen, A., Dideberg, O., Jamin, M., and Dessen, A. (2008). Trapping of an acyl-enzyme intermediate in a penicillin-binding protein (PBP)-catalyzed reaction. J. Mol. Biol. *376*, 405–413.

Magnet, S., Bellais, S., Dubost, L., Fourgeaud, M., Mainardi, J.L., Petit-Frère, S., Marie, A., Mengin-Lecreulx, D., Arthur, M., and Gutmann, L. (2007). Identification of the L,D-transpeptidases responsible for attachment of the Braun lipoprotein to *Escherichia coli* peptidoglycan. J. Bacteriol. *189*, 3927–3931.

Mainardi, J.L., Legrand, R., Arthur, M., Schoot, B., van Heijenoort, J., and Gutmann, L. (2000). Novel mechanism of β-lactam resistance due to bypass of DD-transpeptidation in *Enterococcus faecium*. J. Biol. Chem. *275*, 16490–16496.

Mainardi, J.L., Fourgeaud, M., Hugonnet, J.E., Dubost, L., Brouard, J.P., Ouazzani, J., Rice, L.B., Gutmann, L., and Arthur, M. (2005). A novel peptidoglycan cross-linking enzyme for a β-lactam-resistant transpeptidation pathway. J. Biol. Chem. *280*, 38146–38152.

Mainardi, J.L., Hugonnet, J.E., Rusconi, F., Fourgeaud, M., Dubost, L., Moumi, A.N., Delfosse, V., Mayer, C., Gutmann, L., Rice, L.B., and Arthur, M. (2007). Unexpected inhibition of peptidoglycan LD-transpeptidase from Enterococcus faecium by the β-lactam imipenem. J. Biol. Chem. *282*, 30414–30422.

Mainardi, J.L., Villet, R., Bugg, T.D., Mayer, C., and Arthur, M. (2008). Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. FEMS Microbiol. Rev. *32*, 386–408.

Mittermaier, A.K., and Kay, L.E. (2009). Observing biological dynamics at atomic resolution using NMR. Trends Biochem. Sci. *34*, 601–611.

Palmer, A.G., 3rd. (2004). NMR characterization of the dynamics of biomacromolecules. Chem. Rev. *104*, 3623–3640.

Peltier, J., Courtin, P., El Meouche, I., Lemée, L., Chapot-Chartier, M.P., and Pons, J.L. (2011). *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. J. Biol. Chem. *286*, 29053–29062.

Pelton, J.G., Torchia, D.A., Meadow, N.D., and Roseman, S. (1993). Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. Protein Sci. *2*, 543–558.

Plesniak, L.A., Connelly, G.P., Wakarchuk, W.W., and McIntosh, L.P. (1996). Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: NMR assignments, pH titration, and hydrogen exchange. Protein Sci. *5*, 2319–2328.

Pratt, R.F. (2008). Substrate specificity of bacterial DD-peptidases (penicillin-binding proteins). Cell. Mol. Life Sci. *65*, 2138–2155.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics *23*, 381–382.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. *34*, 93–158.

Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A., and Charlier, P. (2008). The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol. Rev. *32*, 234–258.

Schneck, J.L., Villa, J.P., McDevitt, P., McQueney, M.S., Thrall, S.H., and Meek, T.D. (2008). Chemical mechanism of a cysteine protease, cathepsin C, as revealed by integration of both steady-state and pre-steady-state solvent kinetic isotope effects. Biochemistry *47*, 8697–8710.

Schüttelkopf, A.W., and van Aalten, D.M.F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D Biol. Crystallogr. *60*, 1355–1363.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Shokhen, M., Khazanov, N., and Albeck, A. (2011). The mechanism of papain inhibition by peptidyl aldehydes. Proteins *79*, 975–985.

Tipper, D.J., and Strominger, J.L. (1965). Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. Proc. Natl. Acad. Sci. USA *54*, 1133–1141.

Tsai, C.J., Ma, B.Y., and Nussinov, R. (1999). Folding and binding cascades: shifts in energy landscapes. Proc. Natl. Acad. Sci. USA *96*, 9970–9972.

Van Melckebeke, H., Simorre, J.P., and Brutscher, B. (2004). Amino acid-type edited NMR experiments for methyl-methyl distance measurement in ^{13}C-labeled proteins. J. Am. Chem. Soc. *126*, 9584–9591.

Vollmer, W., and Seligman, S.J. (2010). Architecture of peptidoglycan: more data and more models. Trends Microbiol. *18*, 59–66.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Wang, C.Y., Grey, M.J., and Palmer, A.G., 3rd. (2001). CPMG sequences with enhanced sensitivity to chemical exchange. J. Biomol. NMR *21*, 361–366.

Yu, L., and Fesik, S.W. (1994). pH titration of the histidine residues of cyclophilin and FK506 binding protein in the absence and presence of immunosuppressant ligands. Biochim. Biophys. Acta *1209*, 24–32.

Zapun, A., Contreras-Martel, C., and Vernet, T. (2008). Penicillin-binding proteins and β-lactam resistance. FEMS Microbiol. Rev. *32*, 361–385.

6.4 Supplemental information of the article published in Structure

Supplemental information for:

Dynamics induced by β-lactam antibiotics in the active site of *Bacillus subtilis* L,D-transpeptidase

Lauriane Lecoq, Catherine Bougault, Jean-Emmanuel Hugonnet, Carole Veckerlé, Ombeline Pessey, Michel Arthur and Jean-Pierre Simorre

Inventory of supplemental information

Supplemental Data

Figure S1, related to Figure 2: Superposition of the [^{1}H,^{15}N]-HSQC spectra of Ldt_{Bs} and Ldt_{Bs} C142A

Figure S2, related to Figure 4: Plot of chemical shifts as a function of pH for different nuclei of His122

Figure S3, related to Figure 5: Chemical shift perturbations induced by non-covalent binding of ampicillin and imipenem on Ldt_{Bs} and Ldt_{Bs} C142A

Figure S4, related to Figure 5: Thermodynamics of the interaction between Ldt_{Bs} and Ldt_{Bs} C142A with ampicillin and imipenem, and kinetics of Ldt_{Bs} acylation by imipenem

Figure S5, related to Figure 6: Impact of covalent modification of Ldt_{Bs} by imipenem and DTNB

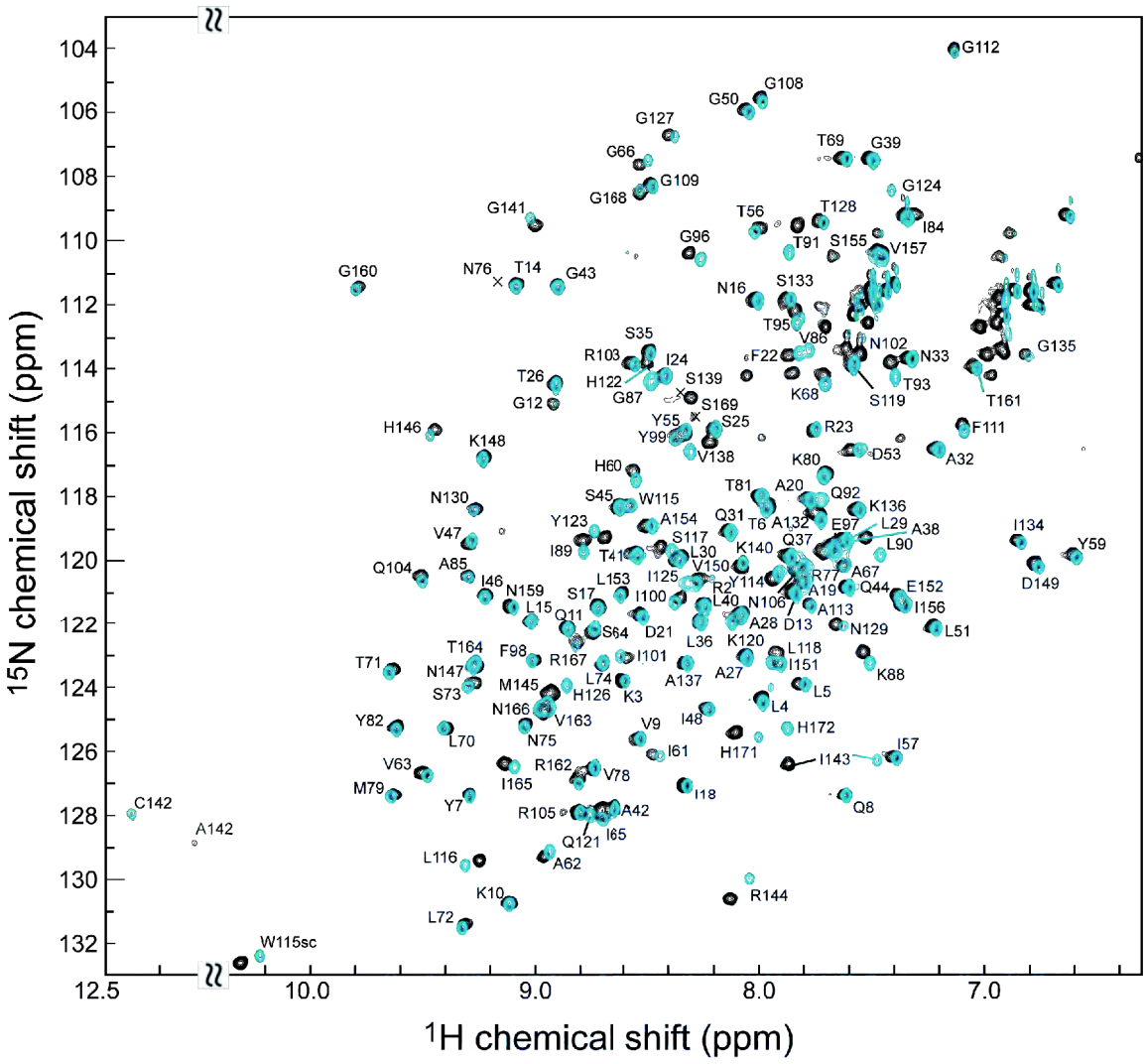
Figure S6, related to Figure 7: Relaxation-dispersion exchange rates for regions 1 to 3 in the three forms of Ldt_{Bs}

Table S1, related to Figure 7: Relaxation data for Ldt_{Bs} and for the imipenem and TNB adducts

Table S2, related to Figure 3: Thermodynamic constants for the protonation and tautomerization of histidines in Ldt_{Bs} at 25 °C and pH 6.5

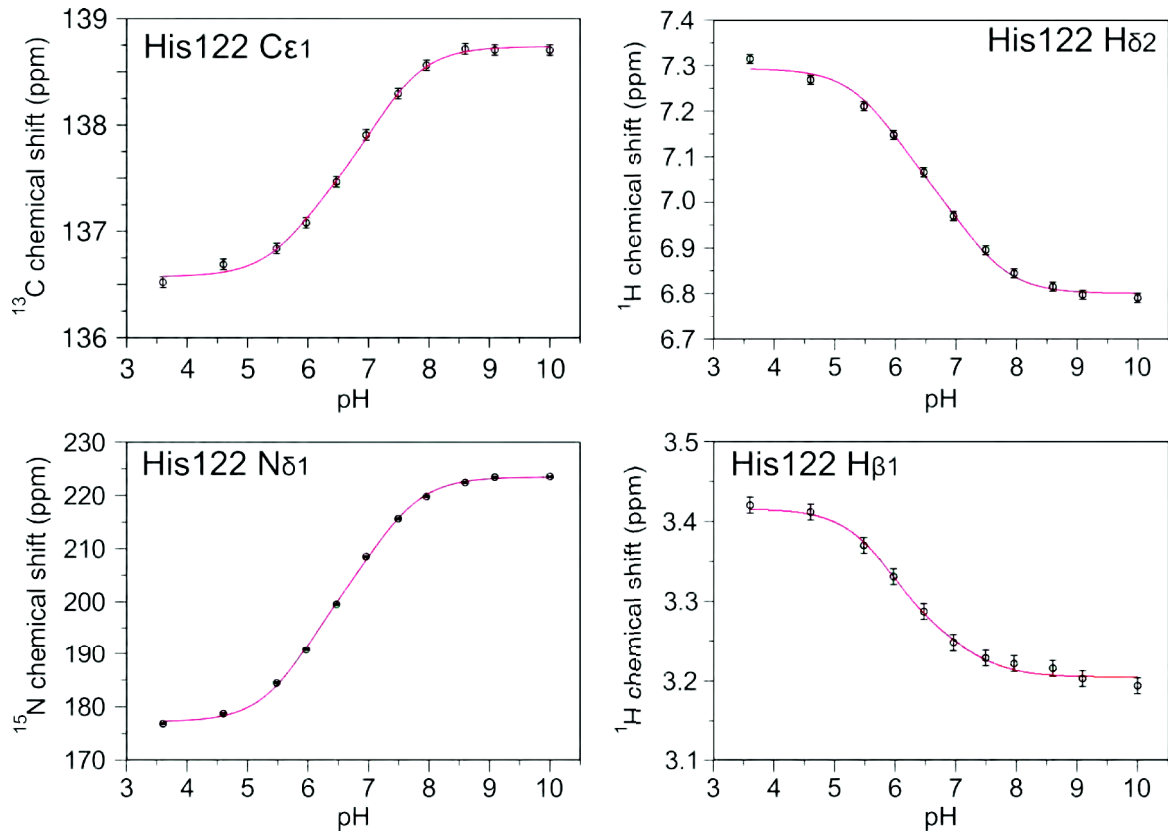
Supplemental Experimental Procedures

Supplemental References



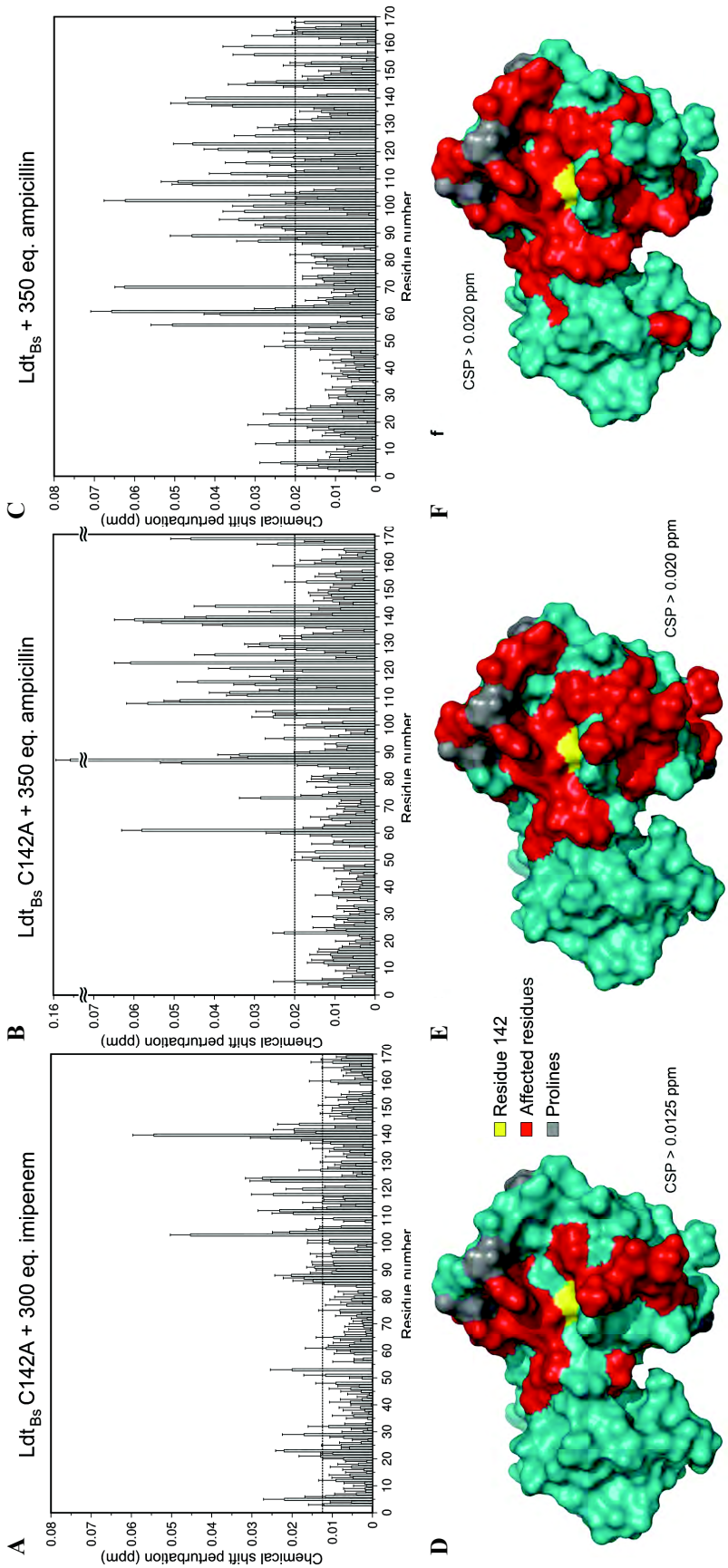
Supplementary Figure 1, related to Figure 2: Superposition of [^{1}H,^{15}N]-HSQC spectra of Ldt_{Bs} and Ldt_{Bs} C142A

Superposition of the 2D [^{1}H,^{15}N]-HSQC spectra of Ldt_{Bs} (blue) and Ldt_{Bs} C142A (black) at 25 °C in 12.5 mM MES and 150 mM NaCl buffer at pH 6.5, recorded at 600 MHz proton frequency. Assignments have been reported on the spectrum with the exception of Gln and Asn side chains (unassigned resonances in the upper right part of the spectrum). Crosses indicate the position of resonances of lower intensity. The side-chain Nε1-Hε1 correlation of the single tryptophane (W115) in the protein sequence is labeled as W115sc. Minor effects were only seen in residues surrounding position 142 indicating that the C142A substitution did not modify the conformation of the protein.



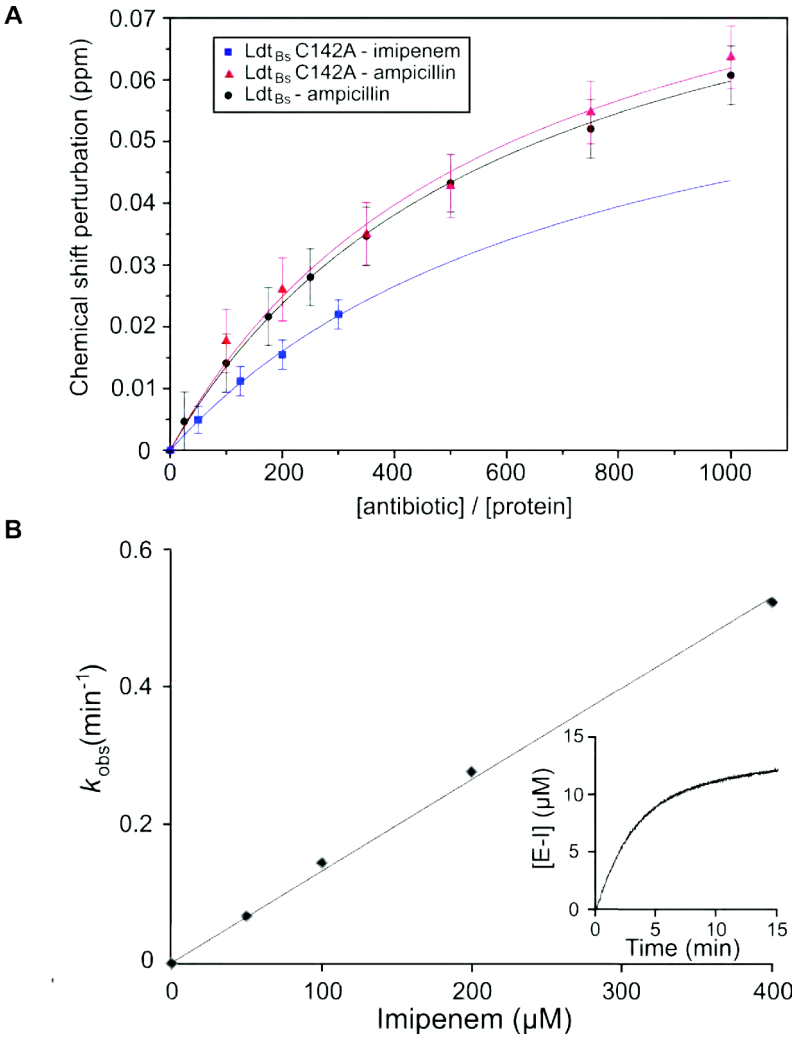
Supplementary Figure 2, related to Figure 4: Plot of chemical shifts as a function of pH for different nuclei of His122

Experimental data points, pictured here for four nuclei of His122, were fitted with a single pKa value (5.91 ± 0.05) for 11 nuclei and the curves deduced from least-square fitting are shown in red. Experimental errors bars are reported on all plots, but stand in some cases (His122 Nδ1, especially) within the symbol width.



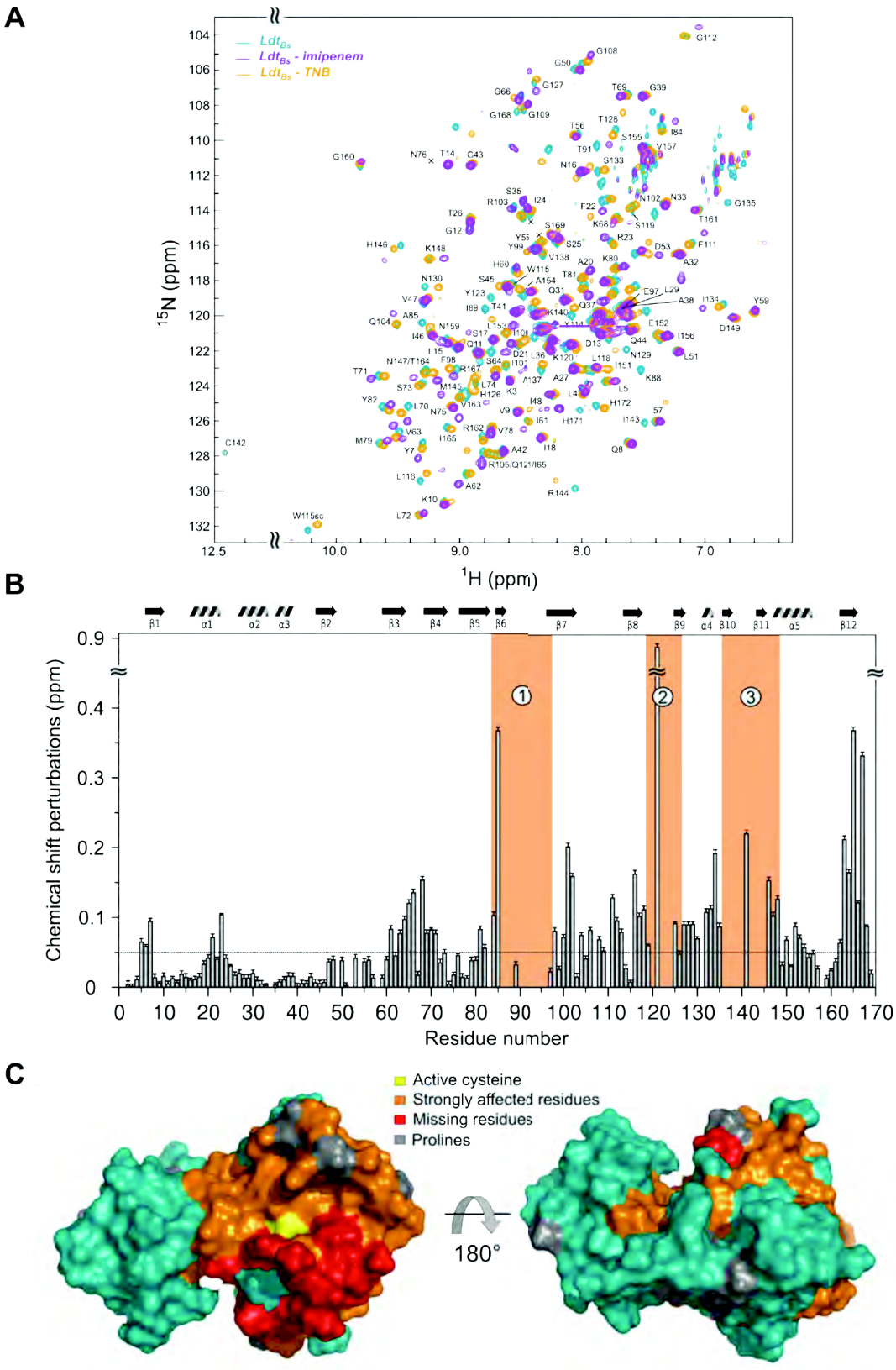
Supplementary Figure 3, related to Figure 5: Chemical shift perturbations induced by non-covalent binding of ampicillin and imipenem on Ldt$_{Bs}$ and Ldt$_{Bs}$ C142A

^{1}H, ^{15}N-chemical shift perturbations (Equation (2) in Supplemental Experimental procedures) induced by (A) imipenem in Ldt$_{Bs}$ C142A (125 µM) and by ampicillin in (B) Ldt$_{Bs}$ C142A (218 µM) and (C) Ldt$_{Bs}$ (164 µM). Residues showing chemical shift perturbations greater than 0.0125 ppm after addition of 300 equivalents of imipenem to Ldt$_{Bs}$ C142A (D) and greater than 0.02 ppm after addition of 350 equivalents of ampicillin to Ldt$_{Bs}$ C142A (E) and to Ldt$_{Bs}$ (F) are indicated in red on the structure of apo Ldt$_{Bs}$. Residue at position 142 (yellow) was also significantly affected along the titration. Prolines are shown in grey on the structure.



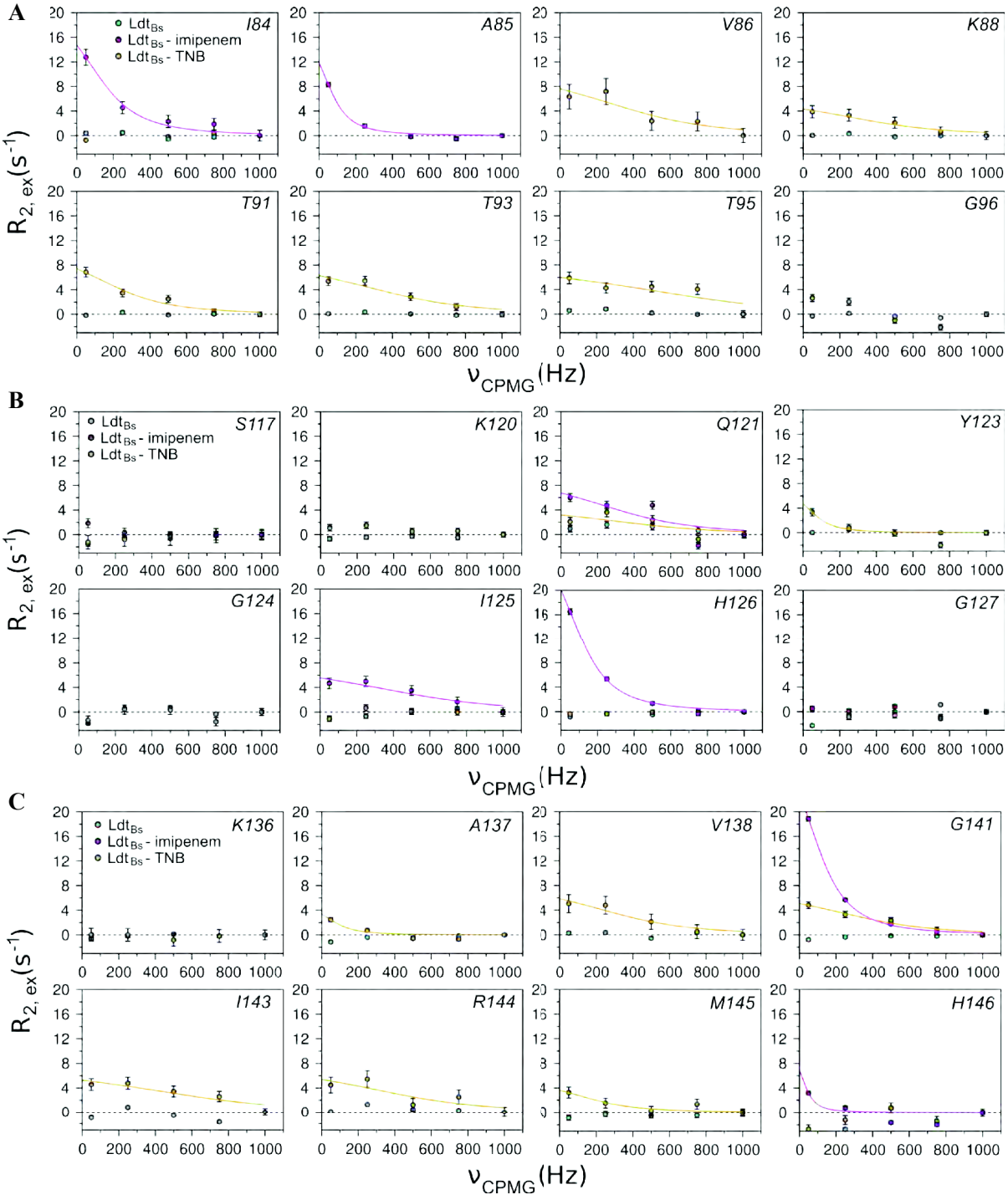
Supplementary Figure 4, related to Figure 5: Thermodynamics of the interaction between Ldt$_{Bs}$ and Ldt$_{Bs}$ C142A with ampicillin and imipenem, and kinetics of Ldt$_{Bs}$ acylation by imipenem

(A) Averaged chemical shift perturbations for all affected residues are reported as a function of the molar ratio between the antibiotic and the protein. The end point of the titration was determined by the solubility limit of the antibiotics. Data were fitted with equation (4) described in Supplemental Experimental Procedures. K_D values of 140 and 110 mM were obtained for binding of ampicillin to Ldt$_{Bs}$ C142A and Ldt$_{Bs}$, respectively. Titration of Ldt$_{Bs}$ C142A by imipenem only provided a lower limit for K_D (> 37.5 mM). (B) Evolution of the pseudo-first order rate constant (k_{obs}) for inactivation of Ldt$_{Bs}$ as a function of imipenem concentration. Inset shows an example of the inactivation kinetics profile of 10 µM Ldt$_{Bs}$ by 100 µM of imipenem, as followed by UV-visible spectroscopy. [E-I] is the concentration of the acylenzyme.


A
Ldt$_{Bs}$
Ldt$_{Bs}$ - imipenem
Ldt$_{Bs}$ - TNB
^{15}N (ppm)
^{1}H (ppm)
B
Chemical shift perturbations (ppm)
Residue number
C
Active cysteine
Strongly affected residues
Missing residues
Prolines
180°

Supplementary Figure 5, related to Figure 6: Impact of covalent modification of Ldt_{Bs} by imipenem and DTNB

(A) Overlay of the 2D [^{1}H,^{15}N]-HSQC NMR spectra of Ldt_{Bs} (blue), the Ldt_{Bs}-imipenem acylenzyme (purple), and the Ldt_{Bs}-2-nitro-5-thiobenzoate (TNB) adduct (orange). All spectra were recorded at 25 °C in 12.5 mM MES and 150 mM NaCl buffer at pH 6.5. (B) ^{1}H and ^{15}N chemical shift perturbation Chemical shift perturbation (Equation (3)) of Ldt_{Bs} resonances upon acylation by imipenem, plotted as a function of the protein sequence after addition of 2.5 equivalents of imipenem. The three regions containing missing resonances or strongly affected residues are highlighted in orange. (C) Surface representation of Ldt_{Bs} showing the position of residues affected by acylation of the enzyme by imipenem. Residues with chemical shift perturbations greater than 0.05 ppm are shown in orange. Residues corresponding to missing resonances are colored in red. Cys142 and proline residues (at positions 34, 49, 52, 54, 58, 83, 94, 107, 110, 131, and 158) are colored in yellow and in grey, respectively.


A
B
C
LdtBs
LdtBs - imipenem
LdtBs - TNB
I84
A85
V86
K88
T91
T93
T95
G96
S117
K120
Q121
Y123
G124
I125
H126
G127
K136
A137
V138
G141
I143
R144
M145
H146
R2, ex(s-1)
νCPMG(Hz)

Supplementary Figure 6, related to Figure 7: Relaxation-dispersion exchange rates for regions 1 to 3 in the three forms of Ldt_{Bs}

Comparison of the measured $R_{2,ex}$ rates in the apo form (blue), the adducts formed with imipenem (purple) and TNB (orange). The 24 residues belong to regions 1 (A), 2 (B), and 3 (C). Increase in the $R_{2,ex}$ contributions evidence conformational exchange. The exchange contribution was fitted as a function of CPMG repetition frequencies with a two-site conformational exchange model (see Supplemental Experimental Procedures). k_{ex} values extracted from the fits were 2 ms^{-1} for the acylenzyme and 6-7 ms^{-1} for the TNB-adduct.

Supplementary Table 1, related to Figure 7: Relaxation data for Ldt_{Bs} and for the imipenem and TNB adducts

Amino acid	Ldt_{Bs} NOE	Ldt_{Bs} R_1 (s^{-1})	Ldt_{Bs} R_2 (s^{-1})	Ldt_{Bs} + imipenem NOE	Ldt_{Bs} + imipenem R_1 (s^{-1})	Ldt_{Bs} + imipenem R_2 (s^{-1})	Ldt_{Bs} + TNB NOE	Ldt_{Bs} + TNB R_1 (s^{-1})	Ldt_{Bs} + TNB R_2 (s^{-1})
R 2	0.140 ± 0.394	1.339 ± 0.183	4.208 ± 0.278	0.014 ± 0.030	1.312 ± 0.100	4.694 ± 0.622	-0.311 ± 0.035	1.367 ± 0.106	6.993 ± 0.749
K 3	0.016 ± 0.010	1.220 ± 0.165	6.358 ± 0.132	0.002 ± 0.009	1.286 ± 0.042	7.003 ± 0.391	0.049 ± 0.007	1.298 ± 0.031	6.370 ± 2.086
L 4	0.522 ± 0.010	1.207 ± 0.035	10.106 ± 0.214	0.510 ± 0.011	1.149 ± 0.037	10.133 ± 0.544	0.457 ± 0.012	1.135 ± 0.024	10.351 ± 0.280
L 5	0.734 ± 0.012	1.282 ± 0.036	11.850 ± 0.372	0.673 ± 0.013	1.222 ± 0.065	13.277 ± 1.349	0.753 ± 0.012	1.188 ± 0.036	11.370 ± 0.365
T 6	0.747 ± 0.010	1.138 ± 0.031	14.370 ± 0.549	0.754 ± 0.012	1.058 ± 0.046	16.524 ± 1.526	0.800 ± 0.010	1.079 ± 0.027	14.397 ± 0.505
Y 7	0.852 ± 0.010	1.072 ± 0.043	13.980 ± 0.771	0.744 ± 0.011	0.988 ± 0.069	17.995 ± 2.348	0.819 ± 0.014	1.044 ± 0.040	15.472 ± 0.720
Q 8	0.753 ± 0.010	1.005 ± 0.021	13.878 ± 0.284	0.759 ± 0.009	0.940 ± 0.030	16.105 ± 0.943	0.746 ± 0.012	0.960 ± 0.019	14.770 ± 0.351
V 9	0.845 ± 0.011	1.110 ± 0.035	13.553 ± 0.377	0.786 ± 0.011	1.060 ± 0.046	15.447 ± 1.177	0.724 ± 0.009	1.088 ± 0.032	13.021 ± 0.489
K 10	0.806 ± 0.009	1.228 ± 0.042	13.891 ± 0.522	0.783 ± 0.010	1.197 ± 0.060	15.051 ± 1.391	0.737 ± 0.009	1.183 ± 0.036	14.106 ± 0.584
Q 11	0.714 ± 0.008	1.233 ± 0.044	12.364 ± 0.236	0.728 ± 0.009	1.146 ± 0.032	13.704 ± 0.708			
G 12	0.691 ± 0.016	1.175 ± 0.173		0.659 ± 0.019	0.957 ± 0.112	36.397 ± 13.854	0.749 ± 0.015	1.003 ± 0.078	
D 13	0.793 ± 0.023	1.184 ± 0.018	13.733 ± 0.225	0.768 ± 0.010	1.135 ± 0.023	15.020 ± 0.606	0.741 ± 0.012	1.148 ± 0.015	13.512 ± 0.257
T 14	0.801 ± 0.008	1.094 ± 0.035	14.033 ± 0.508	0.792 ± 0.008	1.035 ± 0.048	15.415 ± 1.330	0.786 ± 0.010	1.044 ± 0.030	13.720 ± 0.520
L 15	0.748 ± 0.009	1.208 ± 0.028	14.207 ± 0.332	0.789 ± 0.008	1.103 ± 0.034	15.765 ± 0.900	0.772 ± 0.018	1.137 ± 0.023	13.865 ± 0.412
N 16	0.800 ± 0.009	1.172 ± 0.020	12.961 ± 0.222				0.786 ± 0.011	1.148 ± 0.018	11.997 ± 0.260
S 17	0.823 ± 0.009	1.171 ± 0.026	15.228 ± 0.396	0.815 ± 0.009	1.119 ± 0.037	17.235 ± 1.229	0.761 ± 0.010	1.164 ± 0.024	13.692 ± 0.427
I 18	0.849 ± 0.010	1.126 ± 0.034	15.180 ± 0.504	0.807 ± 0.010	1.084 ± 0.051	16.458 ± 1.637	0.798 ± 0.009	1.111 ± 0.033	15.268 ± 0.618
A 19	0.843 ± 0.011	1.185 ± 0.025	14.803 ± 0.429	0.790 ± 0.019	1.137 ± 0.031	17.222 ± 0.949	0.822 ± 0.040	1.168 ± 0.024	13.578 ± 0.428
A 20	0.793 ± 0.010	1.155 ± 0.019	14.838 ± 0.308	0.809 ± 0.013	1.103 ± 0.025	16.607 ± 0.776	0.749 ± 0.011	1.115 ± 0.018	14.707 ± 0.307
D 21	0.785 ± 0.009	1.119 ± 0.023	14.527 ± 0.377	0.788 ± 0.010	1.074 ± 0.036	15.788 ± 1.131	0.802 ± 0.013	1.100 ± 0.022	14.530 ± 0.372
F 22	0.791 ± 0.009	1.092 ± 0.032	14.308 ± 0.563	0.767 ± 0.011	1.030 ± 0.055	18.039 ± 2.251	0.769 ± 0.013	1.098 ± 0.039	14.301 ± 0.649
R 23	0.836 ± 0.008	1.260 ± 0.033	13.775 ± 0.508	0.876 ± 0.022	1.148 ± 0.111	20.938 ± 4.691	0.797 ± 0.014	1.177 ± 0.042	14.764 ± 0.687
I 24	0.793 ± 0.010	1.203 ± 0.025	13.271 ± 0.349	0.775 ± 0.012	1.130 ± 0.032	15.317 ± 0.825	0.759 ± 0.009	1.163 ± 0.023	12.884 ± 0.340
S 25	0.767 ± 0.009	1.132 ± 0.023	11.994 ± 0.315	0.731 ± 0.008	1.055 ± 0.034	13.714 ± 0.772	0.783 ± 0.011	1.084 ± 0.021	12.089 ± 0.301
T 26	0.788 ± 0.010	1.273 ± 0.043	12.886 ± 0.332	0.802 ± 0.030	1.211 ± 0.042	15.311 ± 0.968	0.784 ± 0.009	1.244 ± 0.028	12.821 ± 0.768
A 27	0.792 ± 0.009	1.274 ± 0.023	12.584 ± 0.218				0.743 ± 0.009	1.202 ± 0.018	12.366 ± 0.308
A 28	0.818 ± 0.009	1.158 ± 0.021	13.681 ± 0.254	0.817 ± 0.009	1.107 ± 0.027	15.979 ± 0.749	0.741 ± 0.009	1.149 ± 0.019	12.902 ± 0.265
L 29				0.811 ± 0.007	1.125 ± 0.030	15.638 ± 0.730	0.775 ± 0.012	1.177 ± 0.020	13.058 ± 0.276
L 30	0.793 ± 0.006	1.229 ± 0.027	14.277 ± 0.325	0.784 ± 0.015	1.160 ± 0.037	16.037 ± 0.884	0.755 ± 0.013	1.187 ± 0.024	14.601 ± 0.374
Q 31	0.765 ± 0.009	1.147 ± 0.018	14.198 ± 0.215	0.809 ± 0.008	1.091 ± 0.026	16.466 ± 0.683	0.732 ± 0.008	1.110 ± 0.017	14.186 ± 0.262
A 32	0.757 ± 0.010	1.135 ± 0.020	14.293 ± 0.240	0.811 ± 0.013	1.064 ± 0.029	15.702 ± 0.712	0.800 ± 0.014	1.080 ± 0.019	14.149 ± 0.282
N 33	0.760 ± 0.008	1.115 ± 0.018	11.940 ± 0.185	0.761 ± 0.011	1.055 ± 0.023	13.251 ± 0.509	0.722 ± 0.011	1.089 ± 0.017	11.586 ± 0.205
P 34									
S 35	0.601 ± 0.009	1.232 ± 0.099	12.840 ± 0.315	0.752 ± 0.008	1.159 ± 0.042	13.918 ± 0.913	0.735 ± 0.014	1.222 ± 0.034	12.352 ± 2.870
L 36	0.713 ± 0.008	1.268 ± 0.022	12.852 ± 0.216	0.732 ± 0.011	1.186 ± 0.029	14.978 ± 0.722	0.696 ± 0.009	1.216 ± 0.020	11.411 ± 0.233
Q 37	0.620 ± 0.009	1.338 ± 0.017	11.495 ± 0.120	0.615 ± 0.016	1.280 ± 0.020	12.860 ± 0.329	0.594 ± 0.023	1.267 ± 0.014	11.457 ± 0.165
A 38	0.589 ± 0.009	1.235 ± 0.018	10.467 ± 0.107	0.601 ± 0.011	1.182 ± 0.019	11.625 ± 0.299	0.575 ± 0.006	1.177 ± 0.011	10.887 ± 0.178
G 39	0.586 ± 0.009	1.130 ± 0.016	10.160 ± 0.107	0.607 ± 0.010	1.070 ± 0.019	11.875 ± 0.288	0.588 ± 0.009	1.075 ± 0.012	10.617 ± 0.160
L 40	0.517 ± 0.008	1.047 ± 0.015	9.216 ± 0.093	0.515 ± 0.009	0.987 ± 0.015	10.197 ± 0.252	0.489 ± 0.006	1.028 ± 0.011	9.274 ± 0.172
T 41	0.649 ± 0.011	1.178 ± 0.019	11.363 ± 0.163	0.648 ± 0.008	1.094 ± 0.024	12.689 ± 0.464	0.641 ± 0.010	1.145 ± 0.015	11.031 ± 0.199
A 42	0.721 ± 0.010	1.175 ± 0.027	12.650 ± 0.187	0.747 ± 0.008	1.088 ± 0.026	13.884 ± 0.528	0.735 ± 0.010	1.123 ± 0.019	13.851 ± 0.439
G 43	0.757 ± 0.007	1.160 ± 0.033	12.400 ± 0.321	0.743 ± 0.009	1.066 ± 0.043	13.418 ± 0.884	0.722 ± 0.009	1.127 ± 0.028	11.413 ± 0.337
Q 44	0.768 ± 0.032	1.216 ± 0.017	13.237 ± 0.178	0.767 ± 0.010	1.143 ± 0.022	14.966 ± 0.491	0.746 ± 0.013	1.166 ± 0.015	13.238 ± 0.209
S 45	0.669 ± 0.008	1.077 ± 0.017	12.771 ± 0.190				0.711 ± 0.007	1.050 ± 0.014	12.193 ± 0.194
I 46	0.742 ± 0.009	1.083 ± 0.032	14.750 ± 0.429	0.737 ± 0.008	1.030 ± 0.041	17.237 ± 1.258	0.790 ± 0.011	1.054 ± 0.028	14.033 ± 0.455
V 47	0.793 ± 0.011	1.087 ± 0.035	14.624 ± 0.530				0.751 ± 0.009	1.042 ± 0.031	15.275 ± 0.605
I 48	0.842 ± 0.009	1.192 ± 0.043	12.809 ± 0.497	0.800 ± 0.012	1.157 ± 0.054	15.345 ± 1.433	0.728 ± 0.010	1.172 ± 0.034	12.331 ± 0.467
P 49									
G 50	0.703 ± 0.011	1.138 ± 0.040	13.061 ± 0.453	0.713 ± 0.010	1.054 ± 0.049	14.897 ± 1.075	0.758 ± 0.010	1.087 ± 0.029	14.007 ± 0.536
L 51	0.800 ± 0.008	1.070 ± 0.020	13.695 ± 0.264	0.770 ± 0.009	0.991 ± 0.028	15.290 ± 0.810	0.729 ± 0.014	1.054 ± 0.019	13.530 ± 0.297
P 52									
R 53	0.702 ± 0.012	1.175 ± 0.032	11.902 ± 0.383	0.645 ± 0.013	1.045 ± 0.046	13.416 ± 1.165	0.634 ± 0.009	1.076 ± 0.026	12.028 ± 0.350
P 54									
Y 55	0.780 ± 0.008	1.131 ± 0.033	14.946 ± 0.555				0.751 ± 0.018	1.094 ± 0.026	14.639 ± 0.484
T 56	0.690 ± 0.009	1.205 ± 0.044	13.578 ± 0.623	0.724 ± 0.010	1.105 ± 0.057	15.336 ± 1.790	0.725 ± 0.014	1.115 ± 0.035	13.046 ± 0.603

Supplementary Table 1 (continued)

Amino acid	Ldt_{Bs} NOE	Ldt_{Bs} R_1 (s^{-1})	Ldt_{Bs} R_2 (s^{-1})	Ldt_{Bs} + imipenem NOE	Ldt_{Bs} + imipenem R_1 (s^{-1})	Ldt_{Bs} + imipenem R_2 (s^{-1})	Ldt_{Bs} + TNB NOE	Ldt_{Bs} + TNB R_1 (s^{-1})	Ldt_{Bs} + TNB R_2 (s^{-1})
I 57	0.782 ± 0.009	1.183 ± 0.031	12.736 ± 0.394	0.765 ± 0.009	1.097 ± 0.042	14.122 ± 1.053	0.784 ± 0.012	1.110 ± 0.027	12.563 ± 0.384
P 58									
Y 59	0.738 ± 0.008	1.120 ± 0.029	14.52 ± 0.43	0.748 ± 0.015	1.022 ± 0.039	16.73 ± 1.285	0.706 ± 0.011	1.030 ± 0.026	14.72 ± 0.47
H 60	0.823 ± 0.009	1.061 ± 0.035	18.16 ± 0.89	0.789 ± 0.009	1.011 ± 0.063	20.63 ± 2.509	0.875 ± 0.015	1.003 ± 0.032	17.72 ± 0.77
I 61	0.875 ± 0.013	1.083 ± 0.057	17.41 ± 1.46	0.753 ± 0.012	0.996 ± 0.112	19.62 ± 4.216	0.798 ± 0.012	0.967 ± 0.056	16.52 ± 1.22
A 62	0.810 ± 0.012	1.131 ± 0.039	15.26 ± 0.68	0.827 ± 0.013	1.066 ± 0.057	17.72 ± 1.907	0.890 ± 0.011	1.075 ± 0.038	15.75 ± 0.72
V 63	0.792 ± 0.012	1.025 ± 0.038	14.34 ± 0.64	0.816 ± 0.011	1.000 ± 0.061	16.82 ± 2.091	0.845 ± 0.014	1.007 ± 0.042	15.25 ± 0.82
S 64	0.828 ± 0.011	1.172 ± 0.033	13.97 ± 0.64	0.786 ± 0.008	1.057 ± 0.066	14.83 ± 2.367			
I 65	0.810 ± 0.012	1.142 ± 0.044	14.93 ± 0.90				0.833 ± 0.017	1.097 ± 0.048	15.23 ± 1.12
G 66	0.772 ± 0.010	1.235 ± 0.087	15.06 ± 0.92	0.805 ± 0.013	1.140 ± 0.088	18.74 ± 3.414	0.809 ± 0.015	1.192 ± 0.071	14.49 ± 2.58
A 67	0.827 ± 0.010	1.191 ± 0.047	16.08 ± 0.59				0.833 ± 0.016	1.221 ± 0.036	15.96 ± 1.38
K 68	0.815 ± 0.009	1.143 ± 0.027	13.94 ± 0.54	0.749 ± 0.008	1.108 ± 0.053	18.74 ± 2.454	0.784 ± 0.009	1.100 ± 0.031	14.17 ± 0.60
T 69	0.836 ± 0.009	1.193 ± 0.029	15.20 ± 0.49	0.740 ± 0.009	1.106 ± 0.044	19.18 ± 1.623	0.802 ± 0.009	1.151 ± 0.031	16.03 ± 0.57
L 70	0.852 ± 0.008	1.117 ± 0.039	14.98 ± 0.70	0.828 ± 0.012	1.077 ± 0.069	15.65 ± 2.006	0.849 ± 0.014	1.123 ± 0.046	14.01 ± 0.77
T 71	0.837 ± 0.010	1.125 ± 0.035	14.38 ± 0.61	0.802 ± 0.009	1.082 ± 0.067	17.09 ± 2.107	0.852 ± 0.012	1.108 ± 0.039	13.89 ± 0.65
L 72	0.821 ± 0.011	1.069 ± 0.038	15.67 ± 0.88	0.777 ± 0.009	1.021 ± 0.070	18.34 ± 2.578	0.901 ± 0.012	1.011 ± 0.040	15.61 ± 0.90
S 73	0.793 ± 0.010	1.154 ± 0.046	14.42 ± 0.79	0.818 ± 0.010	1.087 ± 0.068	16.21 ± 2.018	0.748 ± 0.011	1.106 ± 0.045	13.52 ± 0.72
L 74	0.791 ± 0.010	1.171 ± 0.031	14.27 ± 0.43	0.831 ± 0.020	1.060 ± 0.042	16.41 ± 1.262	0.804 ± 0.011	1.065 ± 0.029	13.67 ± 0.46
N 75	0.794 ± 0.009	1.201 ± 0.044	15.02 ± 0.58	0.769 ± 0.010	1.133 ± 0.061	16.93 ± 1.603	0.746 ± 0.012	1.123 ± 0.039	13.55 ± 0.63
N 76	0.733 ± 0.574	1.192 ± 0.524		0.863 ± 0.023	1.334 ± 0.634			1.207 ± 0.294	
R 77				0.741 ± 0.013	1.031 ± 0.026	16.57 ± 0.801			
V 78	0.761 ± 0.007	1.194 ± 0.029	12.78 ± 0.28	0.733 ± 0.009	1.089 ± 0.039	15.70 ± 0.928	0.763 ± 0.011	1.106 ± 0.024	12.94 ± 0.42
M 79	0.793 ± 0.010	1.156 ± 0.056	13.97 ± 0.91	0.721 ± 0.009	1.104 ± 0.089	17.72 ± 2.548	0.765 ± 0.012	1.113 ± 0.055	14.76 ± 0.85
K 80	0.740 ± 0.009	1.063 ± 0.021	14.34 ± 0.39	0.656 ± 0.011	1.013 ± 0.033	15.81 ± 0.984	0.716 ± 0.009	1.014 ± 0.021	14.22 ± 0.38
T 81	0.701 ± 0.014	1.100 ± 0.023	13.72 ± 0.38	0.642 ± 0.010	1.024 ± 0.031	15.19 ± 0.896	0.714 ± 0.009	1.062 ± 0.022	14.06 ± 0.47
Y 82	0.803 ± 0.008	1.106 ± 0.038	14.69 ± 0.69	0.788 ± 0.008	1.048 ± 0.059	16.42 ± 1.912	0.805 ± 0.012	1.067 ± 0.038	14.48 ± 0.71
P 83									
I 84	0.730 ± 0.009	1.162 ± 0.034	11.63 ± 0.45	0.906 ± 0.018	1.094 ± 0.097	19.05 ± 3.757	0.856 ± 0.017	1.111 ± 0.044	12.92 ± 0.60
A 85	0.843 ± 0.008	1.210 ± 0.050	14.30 ± 0.92	0.747 ± 0.027	1.197 ± 0.129	20.71 ± 6.767	0.921 ± 0.026	1.196 ± 0.091	10.83 ± 1.11
V 86	0.845 ± 0.016	1.105 ± 0.035	16.05 ± 0.86				0.938 ± 0.038	1.115 ± 0.095	26.92 ± 9.49
G 87									
K 88	0.816 ± 0.010	1.220 ± 0.033	14.01 ± 0.54				0.809 ± 0.022	1.141 ± 0.065	15.68 ± 1.49
I 89	0.788 ± 0.008	1.347 ± 0.041	16.86 ± 0.64	0.690 ± 0.027	1.402 ± 0.432				
L 90	0.810 ± 0.013	1.242 ± 0.033	17.57 ± 0.77						
T 91	0.745 ± 0.008	1.285 ± 0.032	13.37 ± 0.46				0.672 ± 0.014	1.214 ± 0.103	24.23 ± 5.17
Q 92	0.773 ± 0.009	1.357 ± 0.051	12.71 ± 0.46						
T 93	0.606 ± 0.007	1.249 ± 0.025	12.68 ± 0.35				0.569 ± 0.017	1.173 ± 0.092	36.84 ± 9.15
P 94									
T 95	0.734 ± 0.008	1.320 ± 0.027	12.55 ± 0.39				0.546 ± 0.011	1.317 ± 0.111	37.49 ± 9.61
G 96	0.636 ± 0.008	1.368 ± 0.048	11.56 ± 0.44				0.733 ± 0.019	1.274 ± 0.088	17.08 ± 2.54
E 97							0.658 ± 0.014	1.088 ± 0.024	13.00 ± 0.49
F 98	0.788 ± 0.010	1.147 ± 0.037	13.73 ± 0.67	0.747 ± 0.011	1.025 ± 0.074	14.23 ± 2.228	0.792 ± 0.013	1.068 ± 0.036	13.41 ± 0.60
Y 99	0.794 ± 0.051	1.134 ± 0.051	13.41 ± 0.62				0.815 ± 0.012	1.109 ± 0.054	13.74 ± 0.90
I 100	0.815 ± 0.009	1.204 ± 0.036	13.35 ± 0.58				0.835 ± 0.019	1.158 ± 0.042	13.70 ± 0.68
I 101	0.766 ± 0.012	1.259 ± 0.063	14.40 ± 1.06	0.752 ± 0.016	1.172 ± 0.336	16.94 ± 4.200	0.789 ± 0.021	1.179 ± 0.093	15.97 ± 1.77
N 102				0.745 ± 0.013	1.048 ± 0.066	18.21 ± 1.786	0.828 ± 0.012	1.112 ± 0.034	15.39 ± 0.69
R 103	0.805 ± 0.009	1.189 ± 0.033	13.51 ± 0.34	0.843 ± 0.018	1.099 ± 0.062	19.83 ± 2.720	0.716 ± 0.009	1.128 ± 0.032	14.35 ± 0.70
Q 104	0.781 ± 0.010	1.181 ± 0.034	13.37 ± 0.55	0.782 ± 0.010	1.150 ± 0.109	19.17 ± 4.483	0.830 ± 0.017	1.123 ± 0.043	13.54 ± 0.71
R 105	0.723 ± 0.013	1.154 ± 0.055	14.66 ± 0.55				0.694 ± 0.013	1.097 ± 0.035	15.88 ± 1.48
N 106				0.707 ± 0.028	1.196 ± 0.120	15.79 ± 3.398			
P 107									
G 108	0.600 ± 0.008	1.127 ± 0.025	12.40 ± 0.35	0.629 ± 0.011	1.074 ± 0.064	16.83 ± 2.110	0.540 ± 0.010	1.077 ± 0.027	13.45 ± 0.51
G 109	0.539 ± 0.007	1.095 ± 0.049	11.47 ± 0.34	0.593 ± 0.008	1.069 ± 0.047	15.09 ± 1.046	0.584 ± 0.010	1.074 ± 0.028	12.92 ± 1.14
P 110									

Supplementary Table 1 (continued)

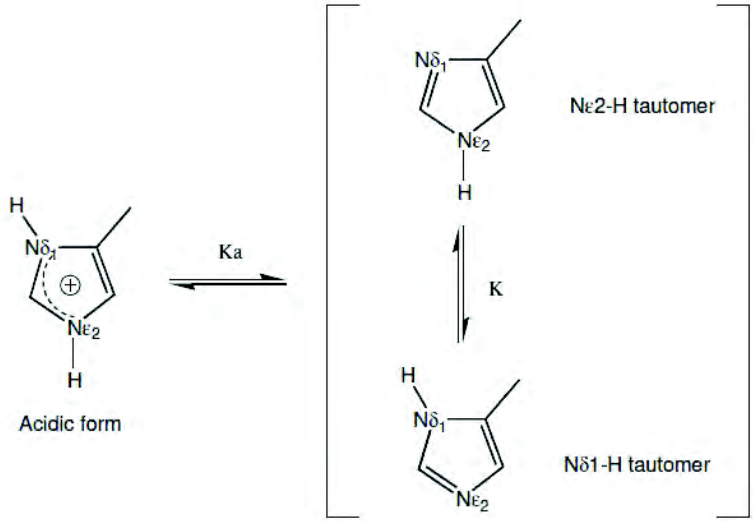
Amino acid	Ldt_{Bs} NOE	Ldt_{Bs} R_1 (s^{-1})	Ldt_{Bs} R_2 (s^{-1})	Ldt_{Bs} + imipenem NOE	Ldt_{Bs} + imipenem R_1 (s^{-1})	Ldt_{Bs} + imipenem R_2 (s^{-1})	Ldt_{Bs} + TNB NOE	Ldt_{Bs} + TNB R_1 (s^{-1})	Ldt_{Bs} + TNB R_2 (s^{-1})
F 111	0.735 ± 0.009	1.166 ± 0.024	12.86 ± 0.41	0.753 ± 0.012	1.048 ± 0.070	16.55 ± 2.623	0.676 ± 0.010	1.090 ± 0.030	13.26 ± 0.50
G 112	0.833 ± 0.013	1.148 ± 0.029	13.82 ± 0.53	0.795 ± 0.012	1.081 ± 0.098	16.95 ± 3.444	0.767 ± 0.009	1.088 ± 0.034	14.05 ± 0.65
A 113	0.798 ± 0.010	1.150 ± 0.034	14.65 ± 0.71	0.844 ± 0.009	1.183 ± 0.138	20.77 ± 9.315	0.816 ± 0.022	1.159 ± 0.035	11.36 ± 0.44
Y 114	0.763 ± 0.011	1.184 ± 0.021	15.20 ± 0.40	0.824 ± 0.045	1.084 ± 0.064	21.14 ± 2.633	0.746 ± 0.009	1.152 ± 0.023	15.29 ± 0.44
W 115	0.792 ± 0.006	1.145 ± 0.034	14.67 ± 0.67				0.772 ± 0.015	1.063 ± 0.043	14.16 ± 0.77
L 116	0.788 ± 0.010	1.121 ± 0.043	14.98 ± 0.75	0.834 ± 0.016	1.059 ± 0.170	25.27 ± 10.208	0.795 ± 0.013	1.036 ± 0.058	16.93 ± 1.15
S 117	0.793 ± 0.009	1.177 ± 0.036	14.45 ± 0.62	0.726 ± 0.012	1.110 ± 0.122	21.53 ± 6.779	0.764 ± 0.014	1.054 ± 0.035	14.33 ± 0.69
L 118	0.794 ± 0.007	1.233 ± 0.041	13.60 ± 0.72				0.768 ± 0.018	1.189 ± 0.048	13.32 ± 0.71
S 119				0.906 ± 0.007	1.027 ± 0.149	17.40 ± 6.552	0.686 ± 0.015	1.081 ± 0.049	13.27 ± 0.81
K 120	0.798 ± 0.006	1.145 ± 0.020	13.67 ± 0.37				0.828 ± 0.012	1.085 ± 0.023	14.03 ± 0.42
Q 121	0.789 ± 0.014	1.118 ± 0.041	15.56 ± 0.79	0.761 ± 0.023			0.946 ± 0.027	1.145 ± 0.108	19.46 ± 2.62
H 122							0.683 ± 0.016	1.090 ± 0.053	12.60 ± 0.84
Y 123	0.815 ± 0.008	1.058 ± 0.029	16.75 ± 0.70				0.708 ± 0.019	1.024 ± 0.055	21.96 ± 2.23
G 124	0.829 ± 0.009	1.096 ± 0.046	14.83 ± 0.89				0.678 ± 0.015	0.963 ± 0.062	15.90 ± 1.10
I 125	0.795 ± 0.016	1.167 ± 0.037	13.64 ± 0.64	0.823 ± 0.015	1.419 ± 0.216	18.03 ± 2.499	0.752 ± 0.018	1.128 ± 0.059	12.86 ± 0.87
H 126	0.833 ± 0.008	1.239 ± 0.033	14.61 ± 0.55	0.823 ± 0.011	1.387 ± 0.292	20.47 ± 4.384	0.841 ± 0.014	1.117 ± 0.035	14.00 ± 0.58
G 127	0.836 ± 0.011	1.138 ± 0.055	14.79 ± 1.07	0.756 ± 0.011	1.004 ± 0.101	19.46 ± 4.243	0.839 ± 0.013	1.033 ± 0.057	15.29 ± 1.10
T 128	0.830 ± 0.009	1.163 ± 0.035	14.26 ± 0.63	0.728 ± 0.009	1.009 ± 0.080	20.99 ± 3.942	0.832 ± 0.016	1.143 ± 0.043	15.81 ± 0.78
N 129	0.811 ± 0.015	1.327 ± 0.083	14.28 ± 1.02	0.782 ± 0.023	1.295 ± 0.441	24.39 ± 59.001	0.771 ± 0.017	1.288 ± 0.090	14.92 ± 4.73
N 130	0.774 ± 0.009	1.184 ± 0.058	15.00 ± 1.07				0.814 ± 0.014	1.186 ± 0.056	15.82 ± 2.07
P 131									
A 132	0.818 ± 0.009	1.249 ± 0.027	13.88 ± 0.35	0.835 ± 0.011	1.200 ± 0.042	15.45 ± 1.228	0.787 ± 0.008	1.202 ± 0.022	14.41 ± 0.60
S 133	0.873 ± 0.009	1.254 ± 0.033	13.50 ± 0.57				0.811 ± 0.013	1.200 ± 0.040	13.10 ± 0.80
I 134	0.802 ± 0.010	1.208 ± 0.034	14.43 ± 0.64	0.778 ± 0.015	1.198 ± 0.093	22.87 ± 4.410	0.763 ± 0.017	1.207 ± 0.046	15.52 ± 0.92
G 135	0.787 ± 0.010	1.162 ± 0.066	16.94 ± 1.63	0.915 ± 0.021	1.031 ± 0.225	18.44 ± 3.529			
K 136	0.840 ± 0.007	1.251 ± 0.027	13.74 ± 0.44				0.762 ± 0.013	1.220 ± 0.031	15.16 ± 0.66
A 137	0.779 ± 0.011	1.149 ± 0.025	13.34 ± 0.42				0.735 ± 0.012	1.123 ± 0.051	16.03 ± 1.36
V 138	0.780 ± 0.012	1.146 ± 0.033	16.61 ± 0.76				0.791 ± 0.017	1.167 ± 0.080	20.54 ± 2.72
S 139									
K 140	0.642 ± 0.014	1.073 ± 0.045	13.67 ± 0.51						
G 141	0.812 ± 0.009	1.183 ± 0.091	13.71 ± 0.89	0.790 ± 0.014	1.034 ± 0.200	20.98 ± 12.553	0.663 ± 0.019	1.144 ± 0.096	22.16 ± 4.03
C 142	0.880 ± 0.020	0.998 ± 0.187	24.98 ± 12.03						
I 143	0.816 ± 0.014	1.163 ± 0.052	17.05 ± 1.16				0.740 ± 0.026	1.078 ± 0.128	19.17 ± 3.53
R 144	0.819 ± 0.009	1.221 ± 0.056	15.49 ± 1.18				0.638 ± 0.026	1.099 ± 0.144	26.01 ± 6.14
M 145							0.804 ± 0.026	1.079 ± 0.099	17.38 ± 2.38
H 146	0.810 ± 0.046	1.224 ± 0.059	15.75 ± 1.31	0.749 ± 0.018	1.126 ± 0.154	19.60 ± 10.089	0.913 ± 0.020	1.151 ± 0.070	17.48 ± 1.60
N 147				0.799 ± 0.017	1.193 ± 0.141	14.58 ± 4.215	0.792 ± 0.010	1.096 ± 0.035	15.37 ± 0.70
K 148	0.839 ± 0.010	1.151 ± 0.026	15.03 ± 0.43	0.787 ± 0.013	1.080 ± 0.109	18.96 ± 3.630	0.818 ± 0.013	1.097 ± 0.026	15.12 ± 0.53
D 149	0.793 ± 0.013	1.213 ± 0.028	14.57 ± 0.49	0.778 ± 0.036	1.149 ± 0.085	17.79 ± 3.343	0.785 ± 0.012	1.149 ± 0.027	13.93 ± 0.45
V 150	0.811 ± 0.021	1.197 ± 0.029	15.79 ± 0.60				0.742 ± 0.012	1.194 ± 0.034	13.09 ± 0.57
I 151	0.846 ± 0.010	1.207 ± 0.033	15.47 ± 0.61	0.808 ± 0.012	1.099 ± 0.057	18.13 ± 1.976	0.813 ± 0.012	1.108 ± 0.036	15.27 ± 0.66
E 152	0.786 ± 0.009	1.139 ± 0.027	15.21 ± 0.49				0.784 ± 0.010	1.091 ± 0.027	15.02 ± 0.50
L 153	0.816 ± 0.009	1.169 ± 0.032	14.92 ± 0.55	0.761 ± 0.030	1.128 ± 0.052	17.11 ± 1.540	0.824 ± 0.014	1.148 ± 0.034	15.10 ± 0.61
A 154	0.782 ± 0.009	1.158 ± 0.025	15.66 ± 0.46	0.798 ± 0.010	1.077 ± 0.045	17.19 ± 1.613	0.847 ± 0.010	1.088 ± 0.025	16.35 ± 0.52
S 155				0.798 ± 0.030	0.988 ± 0.040	16.64 ± 1.222	0.748 ± 0.024	1.000 ± 0.022	13.73 ± 0.44
I 156	0.850 ± 0.010	1.203 ± 0.032	14.30 ± 0.52				0.806 ± 0.010	1.146 ± 0.028	13.42 ± 0.43
V 157				0.679 ± 0.023	1.048 ± 0.053	13.97 ± 1.499	0.763 ± 0.028	1.068 ± 0.028	13.94 ± 0.50
P 158									
N 159	0.763 ± 0.010	1.240 ± 0.046	13.02 ± 0.40	0.813 ± 0.010	1.158 ± 0.048	15.29 ± 1.235	0.804 ± 0.017	1.203 ± 0.031	13.83 ± 0.94
G 160	0.827 ± 0.010	1.237 ± 0.061	12.74 ± 0.81	0.859 ± 0.013	1.113 ± 0.102	14.58 ± 2.484	0.823 ± 0.014	1.171 ± 0.053	13.26 ± 0.75
T 161	0.826 ± 0.011	1.183 ± 0.032	13.73 ± 0.47	0.802 ± 0.010	1.112 ± 0.047	16.87 ± 1.531	0.870 ± 0.014	1.126 ± 0.032	12.97 ± 0.46
R 162	0.836 ± 0.012	1.217 ± 0.049	13.72 ± 0.73	0.742 ± 0.015	1.037 ± 0.340	17.44 ± 3.353	0.817 ± 0.018	1.151 ± 0.051	14.60 ± 0.88
V 163				0.750 ± 0.010	1.079 ± 0.132	19.43 ± 5.374	0.763 ± 0.017	1.093 ± 0.063	14.90 ± 1.05
T 164				0.762 ± 0.009	1.102 ± 0.089	17.02 ± 2.845	0.917 ± 0.015	1.080 ± 0.046	14.58 ± 0.73

Supplementary Table 1 (continued)

Amino acid	Ldt_{Bs} NOE	Ldt_{Bs} R_1 (s^{-1})	Ldt_{Bs} R_2 (s^{-1})	Ldt_{Bs} + imipenem NOE	Ldt_{Bs} + imipenem R_1 (s^{-1})	Ldt_{Bs} + imipenem R_2 (s^{-1})	Ldt_{Bs} + TNB NOE	Ldt_{Bs} + TNB R_1 (s^{-1})	Ldt_{Bs} + TNB R_2 (s^{-1})
I 165	0.798 ± 0.009	1.067 ± 0.037	14.46 ± 0.72	0.842 ± 0.014	1.021 ± 0.113	16.86 ± 4.157	0.804 ± 0.019	1.039 ± 0.061	15.23 ± 1.19
N 166	0.769 ± 0.018	1.149 ± 0.036	12.94 ± 0.53	0.795 ± 0.010	1.094 ± 0.069	17.31 ± 2.143	0.733 ± 0.012	1.120 ± 0.039	14.21 ± 0.68
R 167	0.667 ± 0.009	1.225 ± 0.136	13.72 ± 0.93	0.695 ± 0.016	1.162 ± 0.107	19.79 ± 6.380	0.818 ± 0.023	1.215 ± 0.123	16.30 ± 3.07
G 168	0.484 ± 0.011	1.258 ± 0.143	10.28 ± 0.47	0.574 ± 0.011	1.268 ± 0.082	14.81 ± 1.870	0.570 ± 0.013	1.322 ± 0.061	12.89 ± 0.92
S 169	0.853 ± 0.624	1.232 ± 0.164	5.39 ± 0.65	0.492 ± 0.020	1.263 ± 0.165	10.22 ± 4.678			

The three flexible regions in the covalent adducts are highlighted in orange. Prolines are highlighted in grey. Missing values are due to overlapping peaks L29/Q97, R77/N106, S155/V157, N102/S119, N147/T164 and M145/V163 for Ldt_{Bs}; A27/L118, V47/N130, I100/V150, Q152/I156, I65/R105, A67/Q97, Y55/Y99, S45/W115 and N16/S133 for the Ldt_{Bs}-imipenem adduct; Q11/S64, R77/N106 and G135/NH_2 sidechain for the Ldt_{Bs}-TNB adduct. Missing values are also due low-intensity peaks or missing resonances (G12, N76, S139, S169).

Supplementary Table 2, related to Figure 3: Thermodynamic constants for the protonation and tautomerization of histidines in Ldt_{Bs} at 25 °C and pH 6.5

Histidine residue	pKa[(a)]	K[(b)]	Percentage of the different acido-basic forms for the histidine aromatic ring in Ldt_{Bs} at pH 6.5[(c)]		
					
			Acidic form (%)	Nε2-H tautomer (%)	% Nδ1-H tautomer
His 60	5.76 ± 0.05	0.29 ± 0.13	15 ± 5	65 ± 5	19 ± 5
His 122	5.91 ± 0.05	0.45 ± 0.15	21 ± 5	55 ± 5	25 ± 5
His 146	6.60 ± 0.05	0.54 ± 0.15	56 ± 5	29 ± 5	15 ± 5
His 126	< 3.00	NA	ND	ND	> 99

(a) pKa values are obtained by fitting simultaneously the chemical shifts of Hβa, Hβb, Hδ1, Hε2, Cβ, Cδ1, Cε2, Nδ1, and Nε2 nuclei as a function of pH. His60 and His122 were fitted with equation (6) described in the Supplemental Experimental Procedures section, while His146 was fitted with equation (7).

(b) The percentage of the ε2 tautomer was calculated individually from the observed Nδ1, Nε2 and Cδ2 chemical shifts at basic pH (as $\%\varepsilon_2 = \frac{\delta_{obs} - \delta_{\delta 1 tautomer}}{\delta_{\varepsilon 2 tautomer} - \delta_{\delta 1 tautomer}}$, where $\delta_{\delta 1 tautomer}$ and $\delta_{\varepsilon 2 tautomer}$ stand for the chemical shift values of the corresponding nucleus in the Nδ1-H and Nε2-H tautomer, respectively). Reference values for each tautomer were taken from Day et al., 2003 and Sudmeier et al., 2003 for ^{15}N and ^{13}C, respectively. The three values were then averaged and the constant K was calculated as $K = \frac{1 - \%\varepsilon_2}{\%\varepsilon_2}$.

(c) Percentage of each form calculated from the pKa and K values.

Supplemental Experimental Procedures

Chemicals

Imipenem was kindly provided by Merck. Ampicillin sodium salt and 5-5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) were obtained from Euromedex and Sigma-Aldrich, respectively. Isotopically labeled chemicals for protein preparation, $^{13}C_6H_{12}O_6$ and $^{15}NH_4Cl$, were purchased from Cambridge Isotope laboratories and Eurisotop with 99% enrichment.

Primers

The open reading frame encoding Ldt_{Bs} with a C-terminal His_6-tag was amplified with primers 5'-AACCATGGGGCTGCTTACGTACCAGGTGAAGC-3' and 5'-TTGGATCCCCGGT-TAATCGTGACTCTCGT-3'. The C142A Ldt_{Bs} mutant was amplified with primers 5'-AAGGCTG-TTTCTAAAGGCGCCATTCGCATGCATAAC-AAA-3' and 5'-TTTGTTATGCATGCGAAT-GGCGCCTTTAGAAACAGCCTT-3'. In each case, the PCR product digested with NcoI and BamHI (underlined) was cloned into pET2818 (Magnet et al., 2007).

Protein expression, purification, and modification

E. coli BL21(DE3) cells harboring derivatives of plasmid pET2818 were grown at 37 °C overnight in M9 minimal medium containing 2 $g.l^{-1}$ of ^{13}C-glucose, 1 $g.l^{-1}$ of ^{15}N-ammonium chloride, 100 $\mu g.ml^{-1}$ of ampicillin, and 10 $mg.l^{-1}$ each of thiamine and biotin. Induction was performed at an OD_{600} of 0.8 with 0.5 mM isopropyl-β-D-thiogalactopyranoside for 18 h at 16 °C. Cells were collected by centrifugation, resuspended in 50 mM Tris-HCl (pH 7.5) containing 300 mM NaCl and disrupted by sonication or French press. Cell debris were removed by centrifugation (48,000 *g* at 4°C for 30 min) and the clarified extract was loaded on nickel-nitrilotriacetate agarose (Qiagen, GmbH, Germany). Proteins were eluted with 500 mM imidazole (in 50 mM Tris-HCl, 300 mM NaCl, pH 7.5) and further purified by gel filtration (Superdex 75; GE-Healthcare) in 12.5 mM MES, 150 mM NaCl (pH 6.5). $^{13}C,^{15}N$-Ldt_{Bs} was incubated in 50 mM HEPES (pH 7.2) with 5 equivalent of 5-thio(2-nitrobenzoic acid) (DTNB) (Tawfik, 2009) for 45 minutes to prepare the TNB adduct. $^{13}C,^{15}N$-Ldt_{Bs} was incubated with 2.5 equivalents of imipenem to form the acylenzyme within minutes.

Mass spectrometry analyses

Molecular weight of $^{13}C,^{15}N$-labeled protein samples were confirmed by ESI-TOF mass spectrometry. The protein sample was desalted on-line on a protein macroTrap (C4, Michrom Bioresources) at 300 µl/min for 3 min in a 0.03% trifluoroacetic acid (TFA) aqueous solution and eluted in 66.5% acetonitrile, 33.5% H_2O and 0.03% TFA on a 6210 LC-TOF mass spectrometer

interfaced with a liquid chromatography pump system (Agilent Technologies). Analysis took into account the isotopic enrichment of the starting material for labeled protein samples. Observed (expected) molecular weights: 20,009.88 (20,008.70) Da for Ldt_{Bs}, 19,974.56 (19,983.41) Da for Ldt_{Bs} C142A, and 20,207.15 (20,206.10) Da for the TNB adduct of Ldt_{Bs}.

Acylation kinetics of Ldt_{Bs} by imipenem

Formation of the acylenzyme [E-I] resulting from the addition of imipenem to Ldt_{Bs} was followed in 100 mM sodium-phosphate buffer, pH 6.0 at 20 °C by measuring the decrease in UV absorbance at 299 nm resulting from the opening of the β-lactam ring of imipenem ($\Delta\varepsilon_{299mm}$= -7,100 $M^{-1}.cm^{-1}$) (Triboulet et al., 2011). Fast kinetic experiments were performed with a stopped-flow apparatus RX-2000 (Applied Photophysics) coupled to a Cary 100-Bio spectrophotometer (Agilent). Ldt_{Bs} (10 μM) was incubated with various concentrations of imipenem and the pseudo-first order rate constant (k_{obs}) was deduced by fitting experimental data to the exponential model for each concentration:

$$[E-I] = [E_0]\left(1 - e^{-k_{obs}t}\right) \qquad (1)$$

where [E-I] is the concentration of acylenzyme, $[E_0]$ is the concentration of Ldt_{Bs} at the beginning of the reaction and *t* the time (in minutes).

NMR spectroscopy data collection

Spectra for backbone/sidechain assignments, as well as residual dipolar coupling (RDC) measurements, were recorded at 298 K on a Agilent Direct Drive spectrometer equipped with a cryogenic triple $^{1}H,^{13}C,^{15}N$-resonance probe functioning with a ^{1}H NMR frequency at 600 MHz. 3D ^{15}N-NOESY-HSQC (τ_m = 140 ms), 3D ^{13}C-methyl-selective-NOESY-HSQC (τ_m = 160 ms) and methyl-selective assignment experiments (Van Melckebeke et al., 2004) were recorded using a Agilent Direct Drive 800 MHz spectrometer equipped with a cryogenic triple resonance probe. Finally, 3D aliphatic and aromatic ^{13}C-NOESY-HSQC (τ_m = 120 ms and τ_m = 147 ms, respectively) experiments were collected on a Bruker Avance spectrometer operating at 900 MHz equipped with a TCI cryoprobe and on a Bruker Avance spectrometer operating at 800 MHz equipped with a TCI cryoprobe, respectively. Backbone sequential resonances were assigned using 3D HNCO, HNCACB, CBCA(CO)NH and side-chain resonances were assigned using aliphatic and aromatic CT-[$^{1}H,^{13}C$]-HSQC, 3D (H)CCH-TOCSY and 3D NOESY experiments (Sattler et al., 1999). ^{13}C- and ^{15}N-NOESY-HSQC experiments, as well as a 2D homonuclear [$^{1}H,^{1}H$]-NOESY, were used to extract distance restraints. Relaxation data were collected using pulse sequences from Farrow *et al.* (Farrow et al., 1994) on an Agilent spectrometer operating at 600 MHz frequency. ^{15}N R_1 values were measured from 2D [$^{1}H,^{15}N$]-correlation spectra recorded with relaxation delays of 20, 50, 100, 200, 300, 400, 500, 600, 700, 900, 1200, 1500, 1800 ms and ^{15}N R_1 values were measured with 10,

30, 50, 70, 90, 110, 130, 150, 170, 210 and 250 ms relaxation delays. Data were collected with 2 s recycle delay and two redundant data points were used to evaluate reproducibility and errors. Heteronuclear NOE data were obtained by recording a spectrum with a 3-s saturation period and a 2-s recycle delay and a reference spectrum with no saturation and a 5-s recycle delay period. For relaxation-dispersion experiments, compensation for probe heating, occurring more specifically at fast repetition rates of CPMG pulse trains, was set up. In each series a set of five [^{1}H,^{15}N]-correlation spectra was collected with a recycle delay of 750 ms with CPMG frequencies (ν_{CPMG}) of 50, 250, 500, 750 and 1,000 Hz.

Residual Dipolar Coupling (RDC) data collection

Alignment of the Ldt_{Bs} apo-form and acylenzyme was achieved using polyethyleneglycol /alcohol mixtures (Ruckert and Otting, 2000). A 11.4% (w:w) $C_{12}E_5$-polyethyleneglycol stock solution was prepared in 12.5 mM MES and 150 mM NaCl buffer at pH 6.5 containing 10% D_2O. To 500 µl of this solution, 16.5 µl of *n*-hexanol were added dropwise, while vigorously vortexing, to a final molar $C_{12}E_5$/*n*-hexanol molar ratio of 1.03. The presence of the crystalline phase was confirmed by checking the D_2O splitting using direct ^{2}H NMR detection. The obtained alignment medium (300 µl) was then added to a stock solution of Ldt_{Bs} (200 µl). The mixture was vortexed until homogeneity, yielding a 600 µM Ldt_{Bs} final concentration in a 5.7% (w:w) $C_{12}E_5$/water mixture with a 1.03 $C_{12}E_5$/*n*-hexanol molar ratio and a 24 Hz deuterium splitting at 298K and 600 MHz proton frequency. After RDC data collection on the apo-form of Ldt_{Bs}, 2.5 equivalents of imipenem were added to the solution to generate the acylenzyme. After homogenization of the sample, the measured deuterium splitting was 19 Hz at 298 K and 600 MHz proton frequency, consistent with the splitting measured for the apo-form. The recording of a control [^{1}H,^{15}N]-HSQC showed that 100% of the Ldt_{Bs} protein was converted into the acylated form. 3D ^{1}H,^{15}N,^{13}C-correlation experiments based on HNCO and HN(CO)CA pulse schemes were implemented with the BEST-approach (Schanda et al., 2006) at 600 MHz to allow the measurement of one-bond H_N-N, H_α-C_α and C_α-CO, two-bond H_N-CO, three-bond H_N-C_α and four-bond H_α-H_N couplings (Rasia et al., 2011). A complete set including four 3D experiments was typically collected in 93 hours.

Residual Dipolar Coupling (RDC) data analysis

Spectra were processed within NMRPipe (Delaglio et al., 1995) and resonances were automatically detected and fitted using the NlinLS subroutine within the program. The error on each coupling was calculated according to the following formula:

$$error = \frac{1}{\sqrt{2}}\sqrt{\left(\frac{\Delta x}{SN_x}\right)^2 + \left(\frac{\Delta y}{SN_y}\right)^2} \qquad (2)$$

where  and  are the widths of the peak at half height as determined from a gaussian fit of the lineshape in the direct and indirect dimension, respectively, and SN_x and SN_y are the signal-to-noise ratio in the corresponding dimensions. The error on the resulting dipolar coupling was overestimated as the sum of the individual errors of the isotropic and anisotropic couplings. The alignment tensor was then determined using Module (Dosset et al., 2001). Only the $^1D_{HN\text{-}N}$, $^1D_{H\alpha\text{-}C\alpha}$, and $^1D_{C\alpha\text{-}CO}$, RDC values from rigid parts of the protein, as identified from ^{15}N-relaxation measurements, were used as input data for the calculation of the alignment tensor eigenvalues and eigenvectors. A least-square fit procedure was used to extract the axial and rhombic components of the alignment tensor from experimental RDCs. The two LysM and catalytic domains were first considered independently and gave rise to similar alignment tensors (in agreement with relaxation data suggesting that the two domains presented the same tumbling correlation time). The axial and rhombic components of the alignment tensor converged to $A_a = 9.64 \times 10^{-4}$ and $A_r = 0.65 \times 10^{-4}$ for Ldt_{Bs} and to $A_a = 13.1 \times 10^{-4}$ and $A_r = 3.3 \times 10^{-4}$ for the acylenzyme. These values were inserted in the structure calculation run with ARIA2.3.

NMR titration of wild-type and C142A Ldt_{Bs} with β-lactams

218 and 125 µM solutions of ^{13}C,^{15}N-labeled C142A Ldt_{Bs} in 12.5 mM MES and 150 mM NaCl buffer at pH 6.5 in 10% (v:v) D_2O were titrated with imipenem (to its solubility limit, corresponding to a 300:1 antibiotic:protein molar ratio) and ampicillin (up to a 1,000:1 antibiotic:protein molar ratio). A 164 µM solution of ^{13}C,^{15}N-labeled Ldt_{Bs} in 12.5 mM MES and 150 mM NaCl buffer at pH 6.5 in 10% (v:v) D_2O was titrated with ampicillin (up to a 1000:1 antibiotic:protein molar ratio). ^{1}H and ^{15}N chemical shift changes were monitored in 2D [^{1}H,^{15}N]-HSQC spectra at 298 K along the titration. Chemical shift perturbations were calculated as:

$$\Delta\delta = \sqrt{\left(\Delta\delta_H\right)^2 + \left(\frac{\gamma_N}{\gamma_H}\Delta\delta_N\right)^2} \quad (3)$$

Peaks showing chemical shift changes greater than 0.0125 ppm after addition of 300 molar equivalents of imipenem, and 0.02 ppm after addition of 350 molar equivalents of ampicillin were retained to calculate an average chemical shift perturbation $\Delta\delta_{obs}$ over the corresponding residues. $\Delta\delta_{obs}$ was thus calculated for each antibiotic:protein molar ratio and plotted accordingly. Binding constant could be obtained from a non-linear least-square fit with the following formula that take into account the dilution of the protein sample along the titration:

$$\frac{\Delta\delta_{obs}}{\Delta\delta} = \frac{1}{2}\left[1 + X + K_D\left(\frac{X}{L_0} + \frac{1}{P_0}\right) - \sqrt{\left(1 + X + K_D\left(\frac{X}{L_0} + \frac{1}{P_0}\right)\right)^2 - 4X}\right] \quad (4)$$

where L_0 and P_0 are the concentrations of the protein and the ligand solutions used for the titration and X is calculated as the ratio $\frac{L_0 v_L}{P_0 v_{iP_0}}$ with v_L the added ligand volume and v_{iP_0} the initial volume of the titrated protein.

Determination of cysteine and histidines pKa in Ldt_{Bs}

In order to track the protonation state of His and Cys residues in Ldt_{Bs} variants in different pH conditions, a 1 mM solution of $^{13}C,^{15}N$-Ldt_{Bs} was prepared in 20 mM potassium phosphate buffer containing 10% D_2O. pH values of the solution were then adjusted with 50 mM HCl and NaOH stock solutions and measured on the protein sample with a Mettler Toledo InLab Semi-Micro pH-electrode. For each pH value (3.65, 4.60, 5.48, 5.97, 6.47, 6.96, 7.49, 7.96, 8.60, 9.09, 9.99), cysteine Hβ1, Hβ2, and Cβ, as well as histidine Hβ1, Hβ2, Hδ2, Hε1, Cβ, Cδ2, Cε1, Nδ1 and Nε2 chemical shift information was retrieved. In the case of His126, the chemical shift of Hδ1 was also considered. ^1H and ^{13}C chemical shift information at β position was extracted from a [^1H,^{13}C]-HSQC centered on the aliphatic carbons. ^1H and ^{13}C chemical shift information at δ2 and ε1 positions were retrieved from a [^1H,^{13}C]-CT-HSQC centered in the aromatic region. Hδ2/Hε1 and Nδ1/Nε2 chemical shift information were obtained from a [^1H,^{15}N]-HMQC experiments optimized for the detection of $^2J_{NH}$ couplings in imidazole rings (Day et al., 2003; Pelton et al., 1993). ^1H and ^{15}N chemical shift information at δ1 position of His126 was extracted from a [^1H,^{15}N]-HSQC centered in the imidazole region. Chemical shifts of individual cysteine and histidine nuclei were then plotted as a function of pH and pH-sensitive probes were initially individually analyzed with a simple model derived from the modified form of the Henderson-Hasselbach equation (5) using in-house modified non-linear least squares fitting procedures with Monte-Carlo error analysis from NMRPipe (Delaglio et al., 1995),

$$\delta_{obs} = \frac{\delta_A + \delta_B 10^{n(pH-pKa)}}{1+10^{n(pH-pKa)}} \qquad (5)$$

where δ_{obs}, [illegible] and [illegible] stand for the observed and calculated chemical shift of the acidic and basic species, respectively. The Hill coefficient *n*, when below unity, is the signature of the presence of one or more groups in the vicinity of the residue being observed that titrate in the same pKa region with less than two pKa units difference (Markley, 1975). Analysis of these preliminary results for the different nuclei of a single residue and consideration of proximal titrating residues and/or phosphate buffer, then led us to use different models for the solvent-exposed His60 and His122 (Equation (6)) and for His146 (Equation (7)) that could fit the chemical shift evolution of all nuclei at the same time with a single set of pKa values (Joshi et al., 2001).

$$\delta_{obs} = \frac{\delta_A 10^{-2pH} + \delta_B 10^{-(pH+pKa)} + \delta_C 10^{-(pKa+pKa_1)}}{10^{-2pH} + 10^{-(pH+pKa)} + 10^{-(pKa+pKa_1)}} \qquad (6)$$

$$\delta_{obs} = \frac{\delta_A 10^{-3pH} + \delta_B 10^{-(2pH+pKa_2)} + \delta_C 10^{-(pH+pKa+pKa_2)} + \delta_D 10^{-(pKa+pKa_2+pKa_3)}}{10^{-3pH} + 10^{-(2pH+pKa_2)} + 10^{-(pH+pKa+pKa_2)} + 10^{-(pKa+pKa_2+pKa_3)}} \qquad (7)$$

where *pKa* refers to histidine, pKa_1 refers to the $H_2PO_4^-/HPO_4^{2-}$ couple at 7.21, pKa_2 (fitted to 4.93) refers to Asp149, pKa_3 (fitted to 7.72) refers to Lys148, and δ_A, δ_B, δ_C, δ_D refer to the chemical shift of the considered nucleus in each acido-basic form.

Supplemental References

Day, R.M., Thalhauser, C.J., Sudmeier, J.L., Vincent, M.P., Torchilin, E.V., Sanford, D.G., Bachovchin, C.W. and Bachovchin, W.W. (2003). Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN ' by NMR. Protein Sci. *12*, 794-810.

Joshi, M.D., Sidhu, G., Nielsen, J.E., Brayer, G.D., Withers, S.G., and McIntosh, L.P. (2001). Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. Biochemistry *40*, 10115-10139.

Magnet, S., Arbeloa, A., Mainardi, J.L., Hugonnet, J.E., Fourgeaud, M., Dubost, L., Marie, A., Delfosse, V., Mayer, C., Rice, L.B. *et al.* (2007). Specificity of L,D-transpeptidases from gram-positive bacteria producing different peptidoglycan chemotypes. J. Biol. Chem. *282*, 13151-13159.

Markley, J.L. (1975). Observation of histidine residues in proteins by means of nuclear magnetic resonance spectroscopy Accounts Chem. Res. *8*, 70-80.

Rasia, R.M., Lescop, E., Palatnik, J.F., Boisbouvier, J., and Brutscher, B. (2011). Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. J Biomol NMR *51*, 369-378.

Ruckert, M., and Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc *122*, 7793-7797.

Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006). Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc *128*, 9042-9043.

Sudmeier, J.L., Bradshaw, E.M., Coffman Haddad, K.E., Day, R.M., Thalhauser, C.J., Bullock, P.A. and Bachovchin, W.W. (2003). Identification of histidine tautomers in proteins by 2D $^1H/^{13}C^{\varepsilon 2}$ one-bond correlated NMR. J Am Chem Soc *122*, 8430-8431.

Tawfik, D.S. (2009). Modification of Sulfhydryl Groups with DTNB. In The protein protocols handbook, Handbooks, S.P., ed. (Springer Protocols Handbooks), pp. 873-874.

Triboulet, S., Arthur, M., Mainardi, J.L., Veckerle, C., Dubee, V., Nguekam-Moumi, A., Gutmann, L., Rice, L.B., and Hugonnet, J.E. (2011). Inactivation Kinetics of a New Target of beta-Lactam Antibiotics. J. Biol. Chem. *286*, 22777-22784.

6.5 Conclusions on the carbapenem binding to Ldt_{Bs}

Ligand-to-protein binding can generally be classified as one of the three existing models depicted in figure 6.2: the lock-and-key model (figure 6.2a), the selected-fit model (figure 6.2b), or the induced-fit model (figure 6.2c). Especially, when conformational rearrangement is observed upon binding, it is not always an easy task to distinguish whether a protein binds its ligand in an induced-fit or in a selected-fit mechanism (Hammes *et al.*, 2009; Changeux and Edelstein, 2011; Vogt and Di Cera, 2012). Do the distinct conformations involved already exist, so that the complementary conformation for a ligand would be selected to mediate the reaction (selected-fit), or do particular ligands induce a receptor rearrangement until it adopts an adequate conformation (induced-fit)? The characterization of the three individual states (apo-enzyme, non-covalent imipenem complex, and acylenzyme) obtained by NMR suggest that the case of Ldt_{Bs} is even more complicated and do not really fit in the frame of any of the previously described models. The apoenzyme is globally rigid and is not involved in conformational exchange in the classical µs-to-ms time-scale. The non-covalent complex is structurally and dynamically similar to the apoenzyme and the affinity between the carbapenem ligand and the protein is very weak (> 100 mM) suggesting the absence of predefined ligand-specific cavity. Finally, the acylenzyme encounters significant structural rearrangements, accompanied with an exchange between various conformations at the ms time-scale.

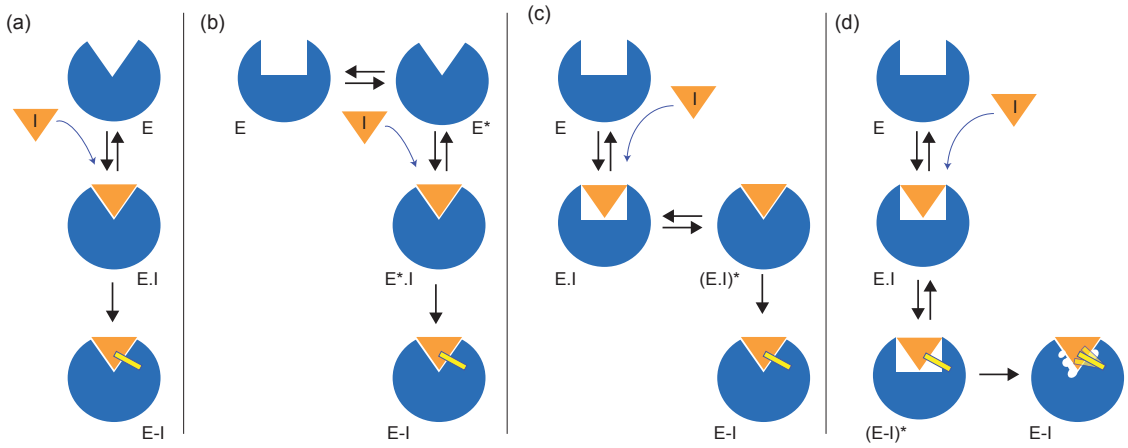


Figure 6.2: Schematic representation of the different ligand-to-protein binding models. a) the lock-and-key model, b) the selected-fit model, c) the induced-fit model, and d) a modified induced-fit model suggested for Ldt_{Bs}. Without the inhibitor I, the conformation E of the enzyme is the ground-state conformation. The star (*) refers to the transition through a so-called "excited" or "invisible" state; the point between E and I (as in E.I, E*.I, (E.I)*) designates a non-covalent Michaelis complex; the dash between E and I (as in E-I and (E-I)*) designates a covalent complex between the enzyme and the inhibitor.

In the lock-and-key model, both the enzyme (E) and the inhibitor (I) possess perfect complementary interfaces (figure 6.2a). The binding does not involve any excited states, instead, the acylenzyme E-I is formed through a simple pathway involving a single non-covalent complex E.I. In Ldt_{Bs}, this model was excluded since conformational changes were observed upon imipenem binding.

The selected-fit model involves a conformational change of the enzyme prior to the ligand binding and postulates for a preexisting equilibrium between the excited (E*) and the ground (E) states (Weikl and von Deuster, 2009) (figure 6.2b). The ligand can then "choose" its preferential enzyme conforma-

tion to form the Michaelis complex E*.I. The excited state E* can usually be detected through NMR relaxation dispersion experiments, that provide dynamic insights into conformational transitions. This model can not account for imipenem-binding to Ldt_{Bs} as no conformational exchange was detected in the apoenzyme. Furthermore, the main structural rearrangements observed on this enzyme occurred only upon covalent binding.

On the contrary, in the induced-fit mechanism, the ligand binds to the ground-state conformation of the protein E, forming a first non-covalent complex E.I. The binding then induces a conformational transition to another state (E.I)*; this step involves in particular a conformational rearrangement at the active site in order to accommodate the inhibitor (figure 6.2c). This model is the closest one that could account for the experimental results obtained on Ldt_{Bs} upon interaction with imipenem. However, the step at which the conformational rearrangement occurs differs drastically. As demonstrated in the article, the non-covalent binding of antibiotics to Ldt_{Bs} induces only minor changes, while the major conformational rearrangement occurs upon formation of the covalent bond between the ligand and the enzyme.

Consequently, for the acylation mechanism of Ldt_{Bs}, a modified-induced-fit model is presented in figure 6.2d. In this model, the inhibitor I first enters within the binding pocket of the enzyme in its ground state E, forming the non-covalent complex E.I whose conformation is very close to the apoenzyme conformation. Then, the covalent binding takes place, generating an excited state (E-I)*, and triggers a conformational change to yield the final stable acylenzyme. As the carbapenem interaction site is quite exposed, the enthalpic contribution to the energy surface is probably moderate and competes with the entropic contribution with the solvent. The resulting acylenzyme E-I is thus characterized by a significant intrinsic flexibility occurring at the µs-ms time-scale.

The results obtained here raise two important questions with respect to the acylation mechanism of Ldt_{Bs} by carbapenems:

First, does the increase in flexibility that was detected upon acylation have a role in the transpeptidation pathway?

Second, does the model of ligand binding suggested here apply to the entire LDt class, can apply to the whole LDt enzyme class, or does the *Bacillus subtilis* enzyme represent a peculiarity?

The first question is particularly challenging to answer, especially because of the extremely low affinity of isolated tripeptides towards LDts (unpublished results from our team). Therefore, the next chapter will focus on the secondly raised-question, through the study of LDts from two other bacterial strains: the LDt from *Enterococcus faecium*, and the LDt from *Mycobacterium tuberculosis*, in order to check whether the conclusions made on Ldt_{Bs} could account for general characteristics of this enzymatic class.

Chapter 7

Towards general mechanistic conclusions on the inhibition of LDts by carbapenem: comparison of Ldt_{Bs} and the LDts from *Enterococcus faecium* and *Mycobacterium tuberculosis*

LDts from *Mycobacterium tuberculosis* have been recognized, in two recent publications in the journals Science (Hugonnet *et al.*, 2009) and Nature Medicine (Gupta *et al.*, 2010), as promising targets in the treatment of tuberculosis in cases of infection by multi- and extensively-drug resistant organisms. Furthermore, LDts are responsible for the transpeptidation of 80 % of the peptidoglycan in stationary phase and dormant forms of the tuberculosis bacillus (Lavollay *et al.*, 2008). Among the five LDts identified in *M. tuberculosis*, two only, Ldt_{Mt1} and Ldt_{Mt2} were shown to be involved in the transpeptidation pathway (unpublished results by the group of Michel Arthur, see 2.3.2.1), thus presenting a particular interest in the context of the present study. As Ldt_{Mt2} ($\sim$ 40 kDa) is found at the upper size limit for structural NMR study, we decided to focus on Ldt_{Mt1} ($\sim$ 30 kDa). *Enterococcus faecium* is the organism in which the L,D-transpeptidase activity has been initially evidenced by the group of Michel Arthur (Mainardi *et al.*, 2000), in relation with increased resistance to ampicillin. In addition to historical reasons, the kinetic of the inhibition of this enzyme by different class of carbapenems has been extensively studied (Triboulet *et al.*, 2011; Dubée *et al.*, 2012a,b). With a size amenable to NMR studies, this enzyme was also retained in the context of the present work.

7.1 Preparation of ^{13}C, ^{15}N-labeled samples for NMR studies

7.1.1 Production and purification of ^{13}C- ^{15}N-labeled Ldt_{Mt1}

The production of a Ldt_{Mt1} sample for NMR studies was particularly difficult. The expression of Ldt_{Mt1} (see figure 7.1a for the protein sequence) in the soluble fraction in M9 minimal medium with $^{15}NH_4Cl$ (1g/L) and ^{13}C-glucose (2 g/L) was quite low and an overall yield of 1 mg of protein per liter of culture was obtained after purification in two steps. SDS-PAGE showed that a large amount of the protein was produced in inclusion bodies, but refolding assays remained unsuccessful to isolate significant amount of protein. These assays were conducted by solubilization of the insoluble fraction in 8 M urea, by purification of the unfolded protein on a Ni-NTA column in urea, and by a final flash refolding in 12.5 mM MES pH 6.5 and purification on a Ni-NTA column in the absence of urea. Furthermore the solubility and stability of the isolated protein was limited: the protein could not be concentrated above 50 µM in 50 mM phosphate buffer containing 300 mM NaCl at pH 6.5, without irreversibly precipitating out of the solution.

Many different conditions were screened by NMR and Thermal Shift Assay (TSA) to try to improve the protein solubility, including variations in buffer, pH, temperature, salt concentration, and the addition of detergents (dodecylphosphocholine, 1-myristoyl-2-hydroxy-sn-glycero-3-(phospho-rac-

1-glycerol), or (3-[3-cholamidopropyl]-dimethylammonio)-1-propanesulfonate), amino acids (Gly, Arg, Glu) or DMSO at various concentrations. Unfortunately, all these tests were not completely satisfying, even if we could conclude that lower salt conditions and higher temperatures favor Ldt_{Mt1} precipitation. From these assays, the highest concentration obtained was 250 µM in 12.5 mM MES at pH 6.5 and without any salt, and the sample remained stable only for a few days.

To improve the expression of a soluble fraction of Ldt_{Mt1}, an alternative technique was tested, consisting in the fusion of Ldt_{Mt1} with the Small Ubiquitin-related MOdifier (SUMO) protein at the N-terminus (figure 7.1b). SUMO is a 100-residue protein that modulates protein structure and function by covalent modification of target proteins in eukaryotes. This protein is actively involved in post-translational modification and participates in many important cellular processes such as transcription, nuclear transport, signal transduction, apoptosis and progression through the cell cycle (Johnson, 2004). The attachment of the SUMO entity to the N-terminus of the target protein facilitates correct protein folding and improves solubility and expression yield in prokaryotic and eukaryotic expression systems (Marblestone *et al.*, 2006; Panavas *et al.*, 2009). Following the expression and purification of the fusion protein, the SUMO-tag can be cleaved by specific proteases via their endopeptidase activity in vitro to generate the desired N-terminus of the released protein partner, here Ldt_{Mt1}. The TEV cleavage site has thus been added between the SUMO and Ldt_{Mt1} constructs.

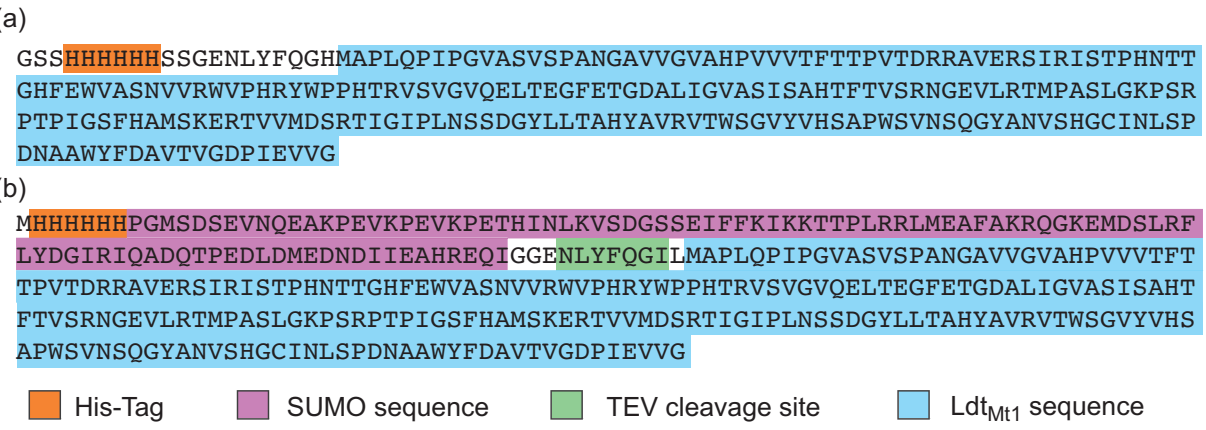


Figure 7.1: Sequence of a) Ldt_{Mt1} and b) the SUMO-Ldt_{Mt1} protein fusion. A 6-His-tag (in orange) is used at the N-terminus to facilitate the first purification step of Ldt_{Mt1} fused to SUMO. The SUMO sequence (in purple) and the Ldt_{Mt1} sequence (in blue) are separated by a cleavage site specific to the TEV protease (in green) in order to easily cleave the SUMO protein and generate the single Ldt_{Mt1} sequence.

The expression of Ldt_{Mt1} fused with SUMO was performed in *E. coli* in M9 medium containing 1 g/L of $^{15}NH_4Cl$ and 4 g/L of ^{13}C-glucose following the standard steps described for Ldt_{fm} (see 7.1.2). The cells were induced overnight at 20 °C, and purified using a first Ni-NTA affinity chromatography and a size exclusion chromatography on a S75 column. SUMO was then cleaved by incubation at room temperature for 1h30 with the TEV protease in a 1:10 TEV:Ldt_{Mt1} (w:w) ratio. The TEV was eliminated from the mixture using a second Ni-NTA affinity chromatography. The SDS-PAGE gels resulting from the two subsequent purification steps are shown in figure 7.2.

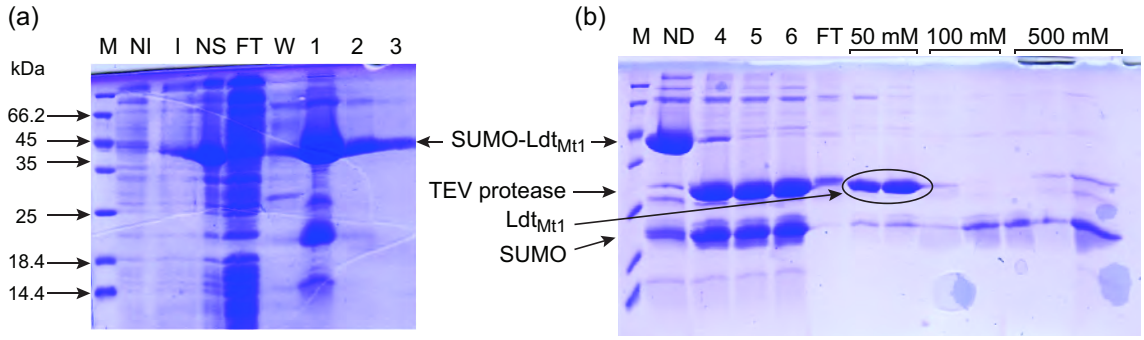


Figure 7.2: SDS-PAGE of Ldt$_{Mt1}$ fused with SUMO, a) prior and b) after cleavage by TEV protease. M : molecular weight marker; NI : Non-induced fraction; I : Culture after induction; NS : insoluble fraction; FT: Flow through; W : Ni-NTA washing step (10 mM imidazole); 1 to 3: Ni-NTA elution fractions (500 mM imidazole); ND : protein before cleavage; 4 to 6 (SUMO-Ldt$_{Mt1}$ with TEV protease loaded on Ni-NTA): 30 min, 1 h, and 1h30 after mixture, respectively; 50 mM, 100 mM and 500 mM: Ni-NTA elution fractions of the TEV/SUMO-LdtMt1 mixture after 1h30 with various imidazole concentrations. Gel a) shows that a minor part of fused SUMO-Ldt$_{Mt1}$ is found in the insoluble fraction (NS), but that the major portion is soluble and can be recovered in the first elution fraction using 500 mM imidazole (hole 1). The subsequent cleavage by TEV protease enables to recover pure Ldt$_{Mt1}$ during elution after 1h30 of incubation and following an elution with 50 mM imidazole.

The gel in figure 7.2a reveals that the SUMO-Ldt$_{Mt1}$ protein construct is mainly recovered in the soluble fraction (FT and hole 1), and 7.2b shows that SUMO is efficiently cleaved after 1h30 incubation with the TEV protease, leading to a pure Ldt$_{Mt1}$ after elution with 50 mM of imidazole. The protein production yield reached 6.33 mg of pure Ldt$_{Mt1}$ per liter of culture in ^{15}N-^{13}C-labeled medium. The SUMO strategy allowed to increase the protein production by a factor of 6, thus proving particularly efficient for the expression and purification of Ldt$_{Mt1}$.

The last step of the protein sample preparation consisted in a dialysis against a 12.5 mM sodium acetate buffer at pH 5.0, to eliminate contaminants from the protein solution. This step is believed to be key in the protein stability and solubility. The protein appears to be stable for at least a month in these low-pH conditions at room temperature and the concentration can be brought up to 1 mM. Samples can be lyophilized or frozen in these conditions for storage. The ^{1}H-^{15}N-correlation spectrum of a 1 mM sample of Ldt$_{Mt1}$ produced with the "SUMO-fusion approach" and conditioned in the optimal pH 5.0 buffer is shown in blue in figure 7.3. This spectrum collected in 10 min is overlaid with a spectrum collected on a Ldt$_{Mt1}$ sample produced with the initial approach. This comparison highlights the significant progresses made to render the tuberculosis sample amenable to NMR structural studies.

The protein production experiments presented in this section were performed by Isabel Ayala and Karine Giandoreggio in our group. The final sample conditions found to reach high Ldt$_{Mt1}$ concentrations were obtained very recently. If these results show that the NMR study on this L,D-transpeptidase is very promising, it will thus not be described further. The present results nevertheless offer nice perspectives for future investigation of the LDt acylation mechanism by carbapenems in a complementary approach to what was pursued on Ldt$_{Bs}$ and the work on Ldt$_{fm}$, described in the next sections.

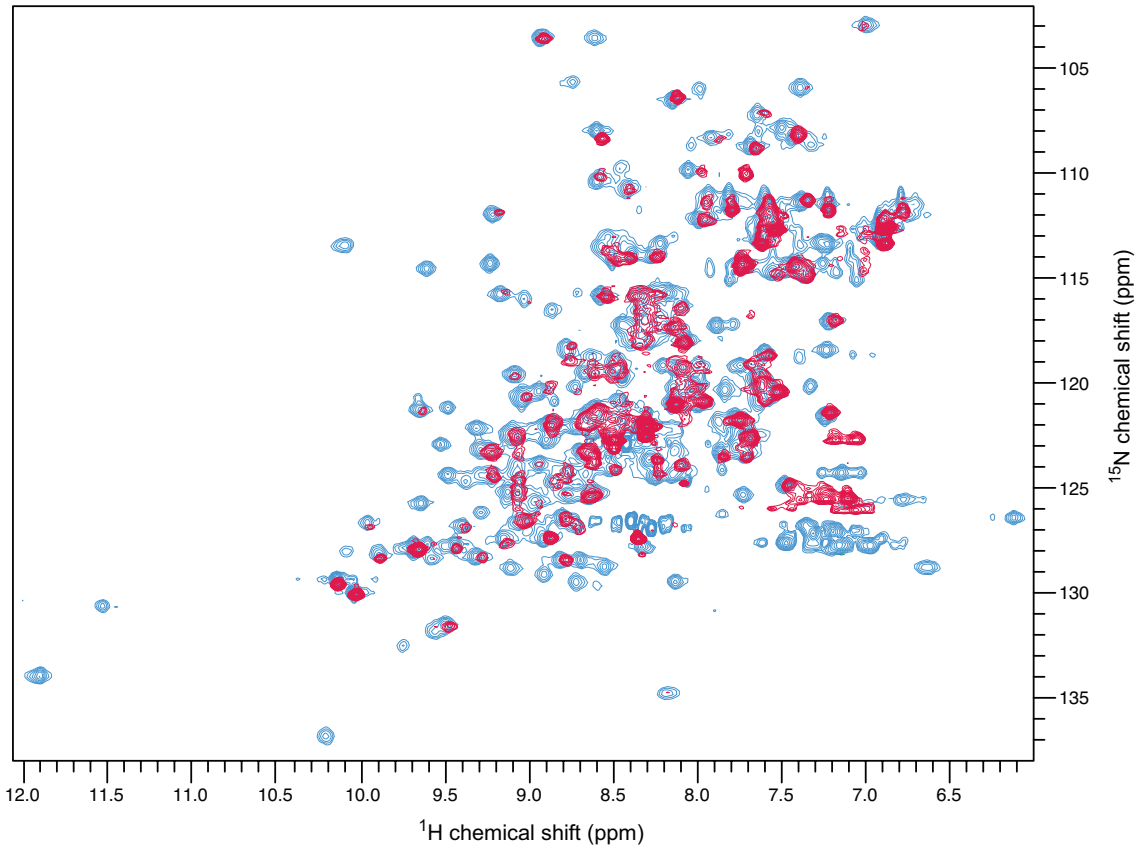


Figure 7.3: ^{1}H-^{15}N-correlation spectra recorded on Ldt_{Mt1} at 25 °C. In red, a ^{1}H-^{15}N-HMQC SOFAST recorded in 2 hours on a 200 µM Ldt_{Mt1} sample in 50 mM sodium phosphate buffer, 100 mM NaCl, pH 6.4. This sample corresponds to the protein sequence given in figure 7.1a and was prepared with the standard approach. In blue, a ^{1}H-^{15}N-HSQC recorded in 10 minutes on a 900 µM Ldt_{Mt1} sample in 12.5 mM acetate buffer at pH 5.0. This sample was prepared using the "SUMO-fusion approach" and the cleavage of SUMO with the TEV protease. The blue spectrum reveals a good signal dispersion, a nice signal-to-noise ratio, and a number of cross-peaks in correlation with the sequence (192 detected isolated correlations for 211 expected). Furthermore, the latter conditions yielded to enhanced stability. The two samples were analyzed by electrospray ionization (ESI) mass spectrometry on a 6210 LC-TOF mass spectrometer interfaced with LC pump system (Agilent Technologies) and yielded observed masses in agreement with the expected masses (Observed (expected) mass for ^{15}N-LdtMt1 after SUMO cleavage : 24,402.25 (24,400.33); Observed (expected) mass for the first ^{15}N-LdtMt1 construct : 26,091.71 (26,092.25)).

7.1.2 Production and purification of ^{13}C-^{15}N-labeled Ldt_{fm}

During the search for optimal conditions of for the NMR study of Ldt_{Mt1}, we started the study of the LDt from *E. faecium*, in the conditions previously optimized by the group of Michel Arthur at the Centre de Recherche des Cordeliers in Paris. Of note, the NMR construct of Ldt_{fm} contains only the catalytic domain of the protein (see figure 7.4 for the protein sequence), as the full length enzyme showed exactly the same kinetics towards antibiotics than the catalytic domain alone (unpublished results, team of Michel Arthur). For clarity's sake, the acronym Ldt_{fm} will thus refer in the following sections to the catalytic domain of Ldt_{fm}, unless otherwise noted.

The production and purification of Ldt_{fm} was done by the team of Michel Arthur using the following protocol. ^{13}C-^{15}N-Ldt_{fm} was produced in *E.coli* BL21 (DE3) cells containing the pETTEVΩldt_{fm} plasmid in minimal medium M9, containing ^{13}C-glucose and $^{15}NH_4Cl$ as unique source of carbon and

nitrogen, respectively. Bacterial growth was followed at 37 °C by Optical Density (OD) at 600 nm and induced with 0.5 mM of IsoPropyl-β-D-1-ThioGalactopyranoside (IPTG) at OD ≈ 0.6. Cells were incubated at 16 °C under 180 rpm (revolutions per minute) agitation, then collected by centrifugation (4 °C, 6,000 rpm, 20 min). The supernatant was eliminated and the cell pellet was resuspended in a 100 mM sodium phosphate buffer at pH 6.4 containing 300 mM NaCl. Cells were disrupted by 5 cycles of 30 seconds of phonolyse, separated by 1 min of rest. After centrifugation (4 °C, 13,000 rpm, 30 min), the soluble fraction was purified by Ni-NTA affinity chromatography, incubated for 1 h at 4 °C, centrifuged again (4 °C, 4,000 rpm, 3 min) and the cell pellet was washed with mL of sodium phosphate buffer containing 100 mM NaCl at pH 6.4. The protein of interest was eluted with 500 mM imidazole and further purified by gel filtration on a Superdex 75 26/60.

```
                    340         350         360         370
HHHHH HENLYFQGHM EDTYIEVDLE NQHMWYYKDG KVALETDIVS

          380         390         400         410         420
GKPTTPTPAG VFYVWNKEED ATLKGTNDDG TPYESPVNYW MPIDWTGVGI

          430         440         450         460
HDSDWQPEYG GDLWKTRGSH GCINTPPSVM KELFGMVEKG TPVLVF
```

Figure 7.4: Primary sequence of the NMR construct of Ldt_{fm}. Residues in green are the residues cleaved by the TEV protease. The catalytic cysteine at position 442 is colored in red. Numbers indicate the position of the amino acid in the complete Ldt_{fm} protein sequence. The first three residues are absent in the full-length protein construct and come from the design of the 6-histidine tag with the TEV cleavage site.

Finally, the 6-histidine tag introduced to facilitate the protein purification was cleaved by the TEV protease. This protease recognizes a linear epitope of the general form E-X-X-Y-X-Q-(G/S), with cleavage occurring between Q and G/S. The most commonly used sequence is ENLYFQG. The amino acids in the initial Ldt_{fm} construct that are cleaved by the TEV protease are shown in green in figure 7.4). The weight ratio TEV/ Ldt_{fm} was 1/10 and TEV was incubated with the protein for 1 h at 30 °C, 50 rpm. The Ldt_{fm}-TEV mixture was further purified on Ni-NTA, Ldt_{fm} was eluted with 50 mM imidazole, centrifuged (4 °C, 4,000 rpm, 3 min) and filtered. The Ldt_{fm} protein sample was further concentrated on a PES Amicon with 10 kDa cutoff, frozen and conserved at -20 °C until the start of the NMR study.

7.2 Thermodynamical and structural NMR studies on the Ldt_{fm} apoenzyme

7.2.1 Assignment of NMR resonances

For immediate comparison with Ldt_{Bs}, a 1 mM ^{13}C,^{15}N-Ldt_{fm} sample was initially prepared in 12.5 mM MES buffer, 150 mM NaCl at pH 6.5. However, in these conditions the protein appeared to be unstable within a week of storage at room temperature. This stability issue was avoided by turning back to the protein purification buffer, 100 mM sodium phosphate buffer, 300 mM NaCl at pH 6.4. The important salt concentration was likely to account for the higher stability. It was thus decided to use these latter conditions, despite the decrease of the NMR-experiments sensitivity on cryogenic probes at such a high salt concentration.

The backbone assignment was initially performed on a protein sample in MES buffer with a set of BEST versions (Lescop *et al.*, 2007) of HNCO, HNCACB and HN(CO)CACB, as well as an HADAMAC (Lescop *et al.*, 2008) experiments collected within 65 hours. Taking into account the high resolution of the ^{1}H-^{15}N-HSQC, a simple BEST-HNCO experiment permitted to transfer the backbone assignments from MES to phosphate buffer conditions. The resulting assigned ^{1}H-^{15}N-HSQC Ldt_{fm} spectrum is shown in figure 7.5. All spectra were recorded at 25 °C on a Agilent Direct Drive spectrometer operating at 600 MHz and equipped with a cryogenic triple resonance probe.

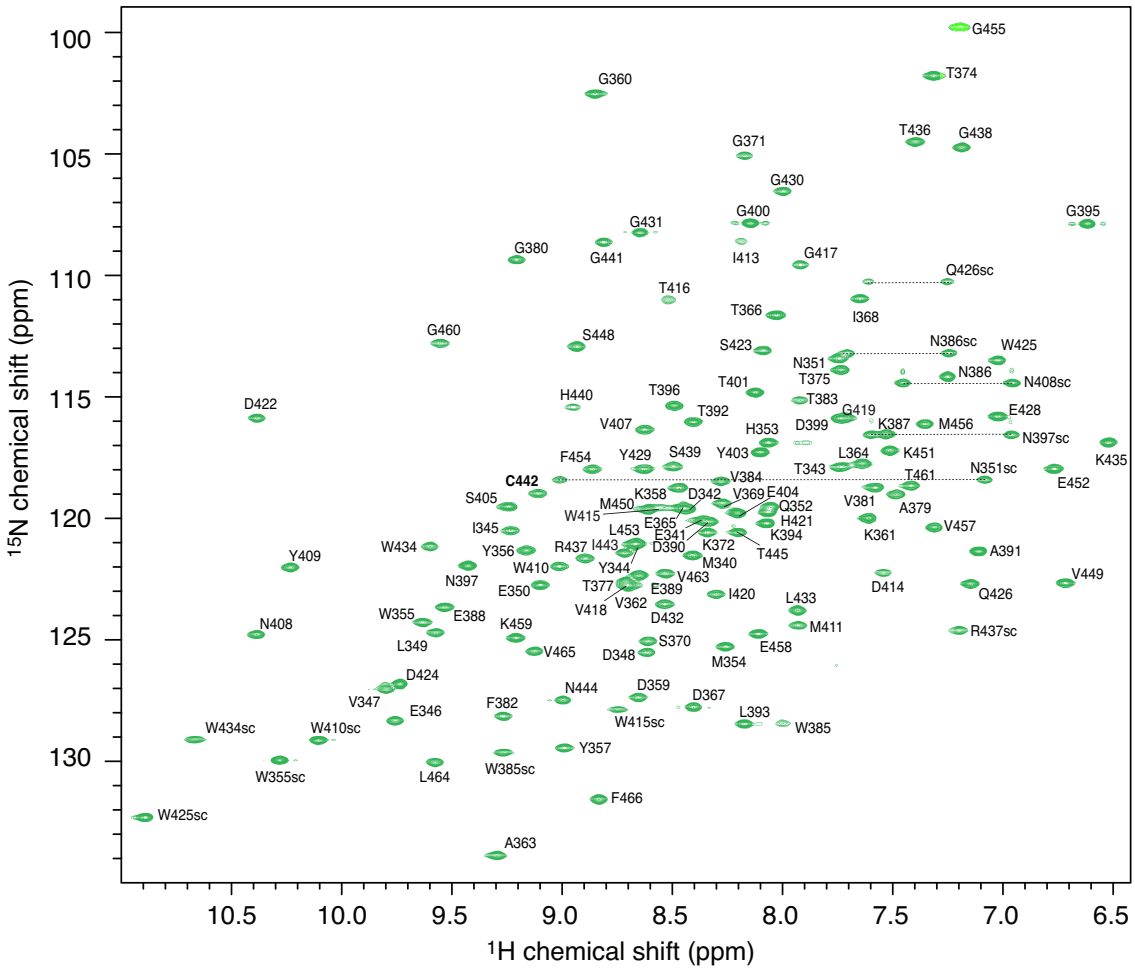


Figure 7.5: ^{1}H-^{15}N-HSQC spectrum of Ldt_{fm} in 100 mM sodium phosphate buffer containing 300 mM NaCl at pH 6.4 and 25 °C. Assignments with amino acid positions in the full-length protein are reported for each correlation. "sc" stands for side-chain amide groups. A horizontal line indicates the side chains of glutamine and asparagine residues.

Side-chain assignments were initially started in MES buffer, using HCCH-TOCSY and ^{15}N-NOESY-HSQC (τ_m = 150 ms) experiments. As Ldt_{fm} contains a high percentage of aromatic residues (15.5 %) among which a large portion at the vicinity of the active site, the following experiments were collected on a spectrometer operating at 600 MHz proton-frequency to extract structural information on His, Phe Tyr and Trp: 2D (HB)CB(CGCD)HD, 2D (HB)CB(CGCDCE)HE, 2D histidine-optimized ^{1}H-^{15}N-HMQC and 3D ^{13}C-NOESY-HSQC (τ_m = 100 ms) centered in the aromatics region. A ^{15}N-NOESY-HSQC (τ_m = 150 ms) collected on a spectrometer functioning at 600 MHz proton-frequency and two ^{13}C-NOESY-

HSQC experiments centered on the aliphatic and aromatic region (τ_m = 120 ms and 100 ms respectively) collected on an Agilent spectrometer functioning at 800 MHz proton-frequency were used to transfer the side-chain assignments from the MES to the phosphate buffer conditions. The same experiments were used to extract distance restraints in the perspective of the structure calculation.

To date, the Ldt_{fm} backbone assignment has been completed to 98.4 % for the non-proline $^{1}H_N$-^{15}N resonances (with the exception of the N-terminal residue), 98.4 % for the $^{13}C\alpha$, 100.0 % for the $^{13}C\beta$ and 85.6 % for the ^{13}C' resonances. In total, 96.0 % of the ^{1}H side-chain resonances were assigned in the apoenzyme. The chemical shifts deposition is under process in the BMRB.

Analysis of Ldt_{fm} chemical shifts using TALOS+ (Shen *et al.*, 2009) is summarized in figure 7.6 and shows that the predicted secondary structure elements are similar to the Ldt_{fm} X-ray secondary structure elements (PDB code: 1ZAT), except that β6 and β7 form a single β-strand in the latter and that the two α-helices are a few residues longer.

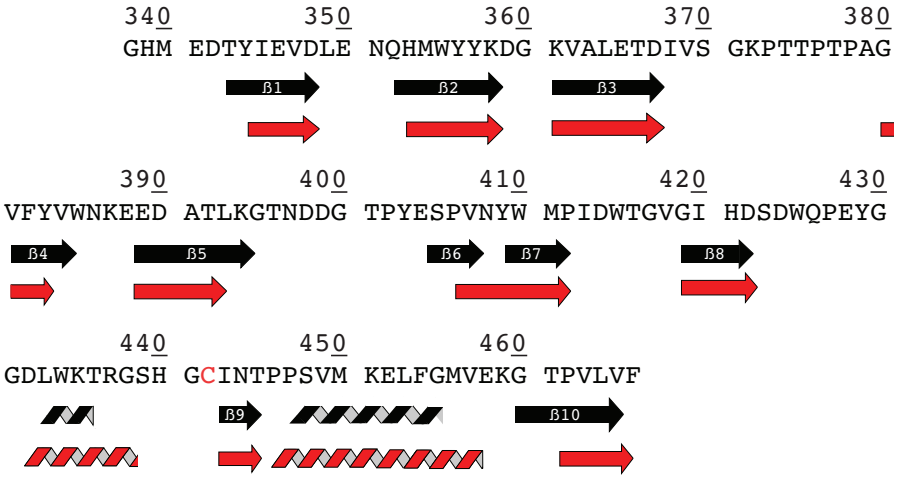


Figure 7.6: Amino acid sequence of Ldt_{fm} with the secondary structure prediction given by TALOS+ from NMR chemical shifts (black) and secondary structure elements from the 1ZAT X-ray structure (red). Helix and β-strand secondary structures are represented by helices and arrows, respectively.

7.2.2 Determination of the protonation state of Ldt_{fm} active-site residues

Similarly to Ldt_{Bs}, the L,D-transpeptidase from *E. faecium* contains a catalytic cysteine (at position 442), and a putative catalytic histidine (at position 421). Due to the presence of a Zn^{2+} ion that probably prevents a correct orientation of the cysteine side-chain at the active site in the crystal structure, it is important to check through our NMR study if the hydrogen bond network identified in Ldt_{Bs} also exists in Ldt_{fm} in the absence of divalent ion. The protonation states of the four histidines and of the cysteine were therefore investigated in Ldt_{fm}, according to the methodology developed in Chapter 4. A sequence alignment is provided in figure 7.7 to visualize the respective positions of histidine and cysteine residues in the two LDts.

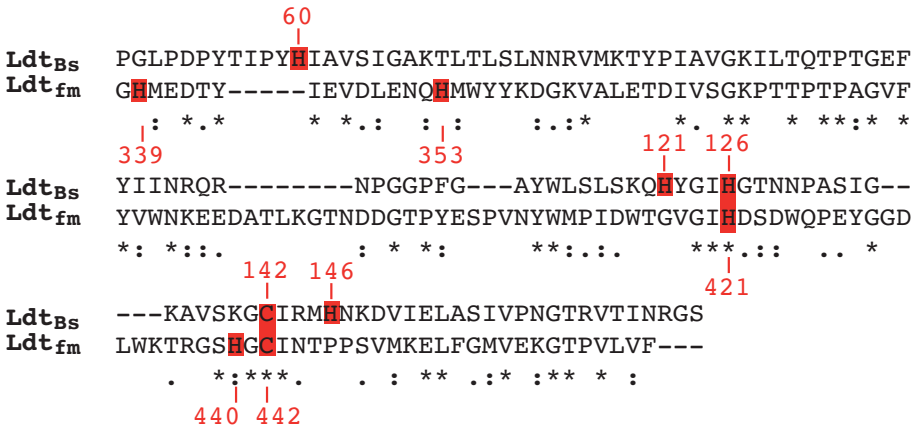


Figure 7.7: Sequence alignment of the catalytic domains of Ldt_{Bs} and Ldt_{fm}. The four histidines of each protein are highlighted in red, as well as the catalytic cysteine.

7.2.2.1 Chemical shifts in the imidazole ring of Ldt_{fm} histidines

Protonation states of histidines were first investigated in the phosphate buffer on the basis of their chemical shifts as presented in section 4.2.4.1. The four histidine (His^{339}, His^{353}, His^{421} and His^{440}) Cϵ1-Hϵ1 and Cδ2-Hδ2 correlations in the ^{1}H-^{13}C-HSQC spectrum were assigned on the basis of the complete backbone assignments and intense intra-residue correlations to the aliphatic Hβ protons in a ^{13}C-NOESY-HSQC centered on the aromatics. Then, Hϵ1 and Hδ2 assignments offered ^{15}N assignments within the imidazole ring through the $^{2}J_{NH}$ or $^{3}J_{NH}$ scalar couplings in a histidine-optimized ^{1}H-^{15}N-HMQC, as sketched in figure 7.8.

At the pH of the study (pH 6.4), the Cδ2 chemical shifts of the four histidines are found around 120 ppm (figure 7.8a), which is characteristic of the presence of either the acidic charged imidazole ring, or the neutral basic Nϵ2-H tautomer. This result is the first indicator that the putative catalytic histidine of Ldt_{fm} (His^{421}) is likely to differ from the one (His^{126}) in Ldt_{Bs}, for which the Cδ2 chemical shift is at 131 ppm, characteristic of the neutral basic Nδ1-H-tautomer. On the contrary to the three other histidines in Ldt_{fm}, the Nδ1 and Nϵ2 nitrogen remain undetectable in the histidine-optimized ^{1}H-^{15}N-HMQC (figure 7.8b), thus preventing to further distinguish the acidic form from the basic Nϵ2 tautomer in His^{421}. This behavior can be explained by the low-sensitivity of the HMQC experiment, or by some exchange broadening, as suggested by the low intensity of the ^{1}H-^{13}C correlations for this residue in comparison to the other histidines (figure 7.8a).

At pH 6.4 in phosphate buffer, the solvent exposed N-terminal His^{339} and the active site His^{440} exist as a mixture of two forms in fast-exchange, a charged acidic form and a neutral basic Nϵ2-H tautomer, as indicated from their characteristic chemical shift pattern. On the contrary, the chemical shift pattern of His^{353} with very close ^{15}N chemical shifts (176.9 and 178.5 ppm) is the signature of a charged acidic form. The Hδ2 chemical shift of this histidine is significantly shifted at 4.38 ppm, and was therefore detected in the ^{1}H-^{15}N-SOFAST-HMQC in figure 7.8). The nitrogen dimension could therefore not be unambiguously assigned as the less intense Nδ1-Hδ2 cross peak was not visible. The localization of His^{353} within the shielding cone of Trp^{355} is likely to account for this unusual Hδ2 chemical shift.

The ^{1}H-^{15}N-HSQC optimized for the detection of $^{1}J_{NH}$ scalar couplings in histidine imidazole rings (left part of figure 7.8) reveals the presence of one peak at 11.76 ppm corresponding to a proton involved

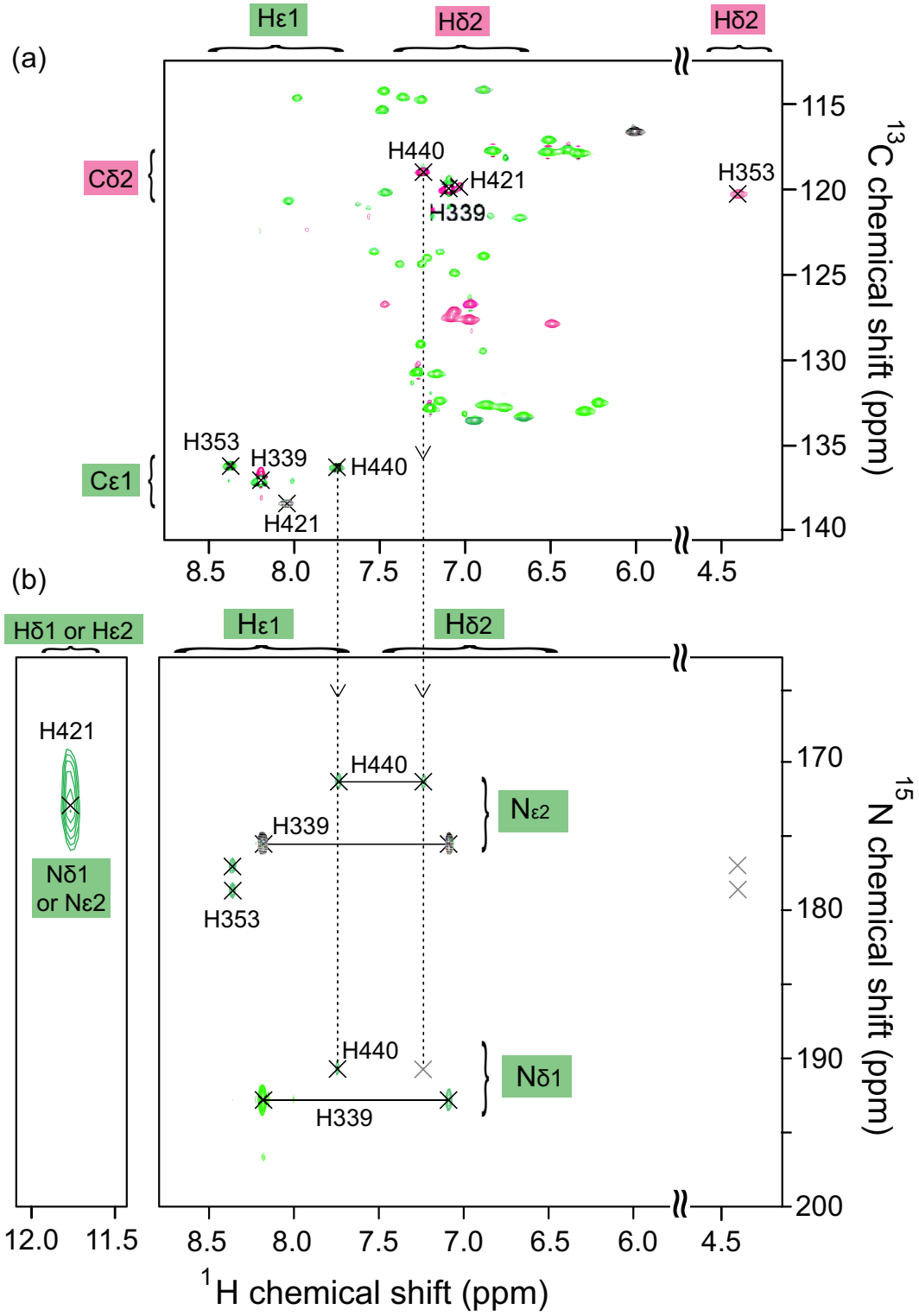


Figure 7.8: 2D NMR spectra collected on Ldt_{fm} in phosphate buffer at pH 6.4 and 25°C to probe the histidine protonation states. a) ^{1}H-^{13}C-CT-HSQC centered on aromatics, displaying the Cϵ1-Hϵ1 and Cδ2-Hδ2 correlations of histidines. Pink and green correlations identify carbons bound to an odd and even number of aromatic carbons, respectively. **b)** ^{1}H-^{15}N-HSQC optimized for the detection of $^{1}J_{NH}$ in histidine imidazole rings (left) and ^{1}H-^{15}N-HMQC optimized for the detection of $^{2}J_{NH}$ and $^{3}J_{NH}$ scalar couplings in histidine imidazole rings (right), used to derive Nδ1 and Nϵ2 chemical shifts. The common ^{1}H dimension enables to correlate the ^{1}H-^{13}C-HSQC with the ^{1}H-^{15}N-HMQC, as shown for His^{440} (dotted lines). The gray crosses correspond to resonances that should appear in the spectrum but are not intense enough because of a small $^{3}J_{NH}$ coupling or because they are out of the frequency window excited in the experiment (His^{353}).

in hydrogen bonding or solvent protected. As the nitrogen dimension is correlated with none of the three assigned histidines, it was naturally proposed as belonging to His^{421}. However, this additional experiment did not provide any additional clue to differentiate between the Nδ1-Hδ1 and the Nϵ2-Hϵ2 correlation. The pH titration was pursued to solve this question, in the hope to detect His^{421} correlations at other pHs.

7.2.2.2 Ldt_{fm} pH titration and pKa determination

pH titration was performed on a 800 µM ^{13}C-^{15}N-labeled Ldt_{fm} sample in 100 mM sodium phosphate buffer containing 300 mM NaCl. Ldt_{fm} was soluble from pH 5.2 to pH 9.41, and precipitated beyond these limits. pH values were adjusted with HCl and NaOH solutions and measured on the protein sample with a Mettler Toledo InLab Semi-Micro pH-electrode. The following experiments were collected on each of the 10 different titration points corresponding to the 5.19, 5.79, 6.23, 6.51, 6.75, 7.12, 7.63, 8.20, 8.81 and 9.41 pH values: a ^{1}H-^{15}N-HSQC, a ^{1}H-^{13}C-HSQC centered on aliphatics, a ^{1}H-^{13}C-Constant Time (CT)-HSQC centered on aromatics, a ^{1}H-^{15}N-HMQC optimized for the detection of $^{2}J_{NH}$ and $^{3}J_{NH}$ scalar couplings in histidine imidazole rings, and a ^{1}H-^{15}N-HSQC centered in the imidazole region to detect the $^{1}J_{NH}$ coupling.

During the titration, His^{339}, His^{353} and His^{440} show a typical chemical shift dependence with pH. A very interesting point observed during this pH titration was the chemical shift of His^{421} Cδ2 (data not shown), that drastically changed between pH 6.51 and pH 6.75: below pH 6.51, it was observed around 120 ppm (characteristic chemical shift for the acidic charged form or the Nϵ2-H tautomer imidazole ring), while above pH 6.75 it was found around 130 ppm (characteristic chemical shift for the Nδ1-H tautomer). This behavior is particularly surprising, and indicates that the Nδ1-tautomeric form is present at least from pH 6.75 to 9.41. The chemical shift of this Nδ1 nucleus is expected between 165 and 175 ppm and could thus consistently correspond to the $^{1}J_{NH}$ correlation shown in the left panel of figure 7.8b that is detected over the whole 5.2-9.5 pH range and that was ambiguously assigned to either Nδ1-Hδ1 or Nϵ2-Hϵ2 of His^{421} at pH 6.4. All these results suggest to propose an assignment of the above mentioned ^{1}H-^{15}N correlation to the Nδ1-Hδ1 correlation of this histidine. Accordingly, His^{421} would evolve from a major charged acidic form below pH 6.5 to a neutral basic Nδ1-H tautomer above this pH. Surprisingly, no Cβ and Hβs chemical shift changes are detected in the ^{1}H-^{13}C-HSQC centered on the aliphatic region, on the contrary to other backbone nuclei. This may due to compensation between a direct acido-basic effect on this histidine and neighboring titrating residues as emphasized in the next paragraph. Alternatively, the chemical shift changes observed on Cδ2 may come from an indirect effect from neighboring titrating residues, leaving the Nδ1-H basic tautomer prevalent on the whole pH-range.

The Cβ and Hβa of Cys^{442} are also not titrated with pH (figure 7.9a and b). However, Cys^{442} N_H, H_N and Hβb nuclei encounter chemical shift changes during the pH titration. It is quite unusual to observe the two phenomena within a same residue. As the Cβ and Hβs are normally affected by the intra-residue protonation state change, the chemical shift variations observed on the other nuclei may be due to an indirect effect coming from a neighboring ionizable residue. To confirm this hypothesis, the pKa determination will be combined with a detailed analysis of the structure.

All titrating nuclei within each histidine and within the cysteine were fitted together in order to get

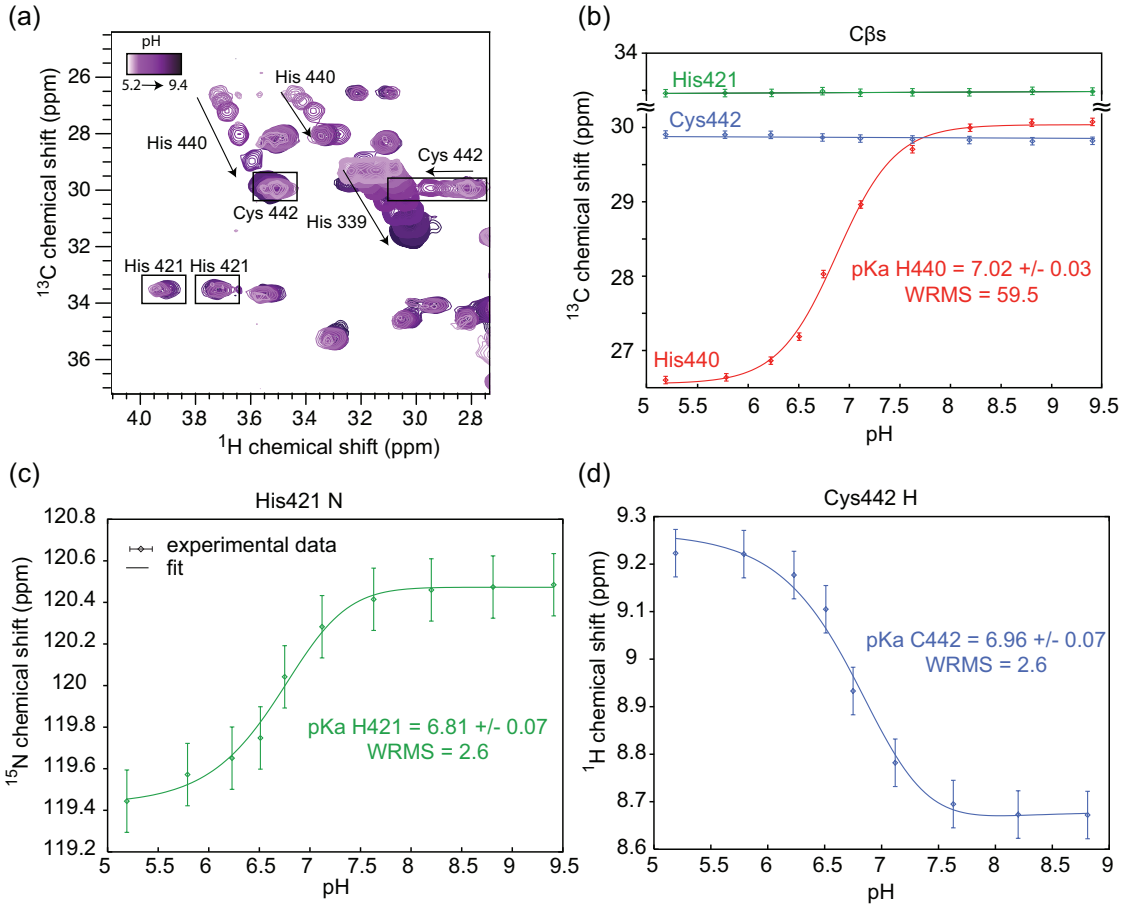


Figure 7.9: pH-titration of Ldt_{fm} followed by NMR. a) ^{1}H-^{13}C-HSQC showing the aliphatic region of histidines and Cys^{442}. Cβ and Hβs of His^{339} and His^{440} are titrating, while β nuclei of His^{421} and Cys^{442} do not exhibit chemical shift changes along the pH titration, except one of the Hβs of Cys^{442}, suggesting an indirect effect. Plot of chemical shifts as a function of pH for **b)** the Cβ nuclei of His^{421}, His^{440} and Cys^{442}; **c)** N_H nucleus of His^{421}; and **d)** H_N nucleus of Cys^{442}. Experimental data points were fitted with equation 4.16 using two pKa values, pKa_1 being the phosphate pKa value fixed at 7.21. The pKa values resulting from collective fit on all available nuclei are very similar for the three residues.

a global pKa value and to reduce uncertainties. His^{353} could not be fitted as properly its pH jump was observed between pH 8 and pH 10. As our experimental data stop at pH 9.41, the plateau at basic pH was not detected, enabling only to estimate a limit value of the pKa, approximately superior to 9.

The three other histidines (His^{339}, His^{421} and His^{440}) were first fitted with a single pKa value with equation 4.14 (page 115). The fits were not satisfying, except for His^{339}, for which a pKa value of 6.85 ± 0.01 was obtained. The two remaining histidines were fitted with 2 pKa values using equation 4.16. 13 nuclei were fitted in the case of His^{440} (N_H, H_N, Cα, Hα, Cβ, Hβa, Hβb, Nδ1, Nϵ2, Hδ2, Hϵ1, Cδ2, and Cϵ1), and 9 nuclei were fitted in the case of His^{421} (N_H, H_N, Cα, Hα, Hδ2, Hϵ1, Cδ2, Cϵ1, and Hδ1). The fitted pKa was 7.02 ± 0.03 for His^{440} (see figure 7.9b for the fit of the Cβ nucleus), and 6.81 ± 0.07 for His^{421} (see figure 7.9c for the fit of the N_H nucleus). In both cases, the second pKa value was found to match with the phosphate pKa (7.21). This is in accordance with the fact that His^{440} is exposed to the solvent, however, this is quite surprising for His^{421} which is embedded within the active site. A pKa of 6.81 for His^{421} also differs drastically from the behavior of the catalytic histidine in Ldt_{Bs}, which did not titrate between 3.5 and 10.

The Cys^{442} nuclei show a similar behavior to that observed for His^{421}: Cβ and Hβa do not shift along the pH titration, while N_H, H_N and Hβb record chemical shift changes. Cα and Hα nuclei could not take part of the analysis as they are covered by the water signal in the aliphatic ^{1}H-^{13}C-HSQC. A Hill coefficient of 1.78 was obtained with equation 4.15 on the three titrating nuclei, , suggesting a cooperative effect between two pKas. The data were therefore subsequently fitted with 2 pKas, one originating from the phosphate and the other was determined at a value of 6.96 $\pm$ 0.07 (see figure 7.9d for the fit of the H_N nucleus).

As the determined-pKa of His^{421} and Cys^{442} are separated from His^{440} pKa by only 0.2 units and 0.06 units, respectively, it strongly suggests that the nuclei that are shifting along the pH titration are in fact influenced by the protonation state change of His^{440}. A structural analysis of the active site on a preliminary NMR structure revealed the presence of His^{440} in the vicinity of His^{421} and Cys^{442}. All the nuclei that are shifting along the pH titration are found at less than 10 Å from the Nδ1 or Nϵ2 nucleus of His^{440} (figure 7.10).

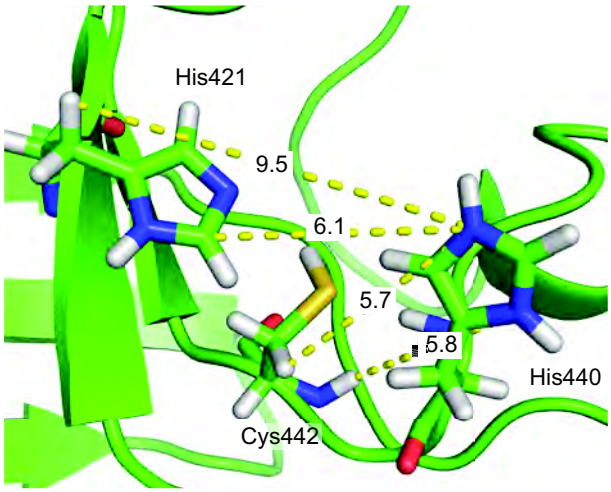


Figure 7.10: Zoom on the active site of Ldt_{fm} showing the proximity of His^{440} with His^{421} and Cys^{442}. The structure shown is a preliminary NMR structure. Orientation of His^{440} is different than in the X-ray structure, as it points into the active site instead of pointing out. Some of the relevant distances between nitrogen atoms of the imidazole ring of His^{440} and different nuclei from His^{421} and Cys^{442} are displayed in Å.

To verify that His^{440} could be the origin of the unusual chemical shifts changes observed for the two catalytic-site residues, the expression of Ldt_{fm}-H440A mutant is in progress. A pH titration will be performed on this mutant and will reveal the corresponding pKa values of His^{421} and Cys^{442} without the influence of a neighboring titrating residue, or the absence of protonation state changes for these residues. If the latter hypothesis is confirmed, the inconsistencies between the Cδ2 and Nδ1 chemical shifts observed in His^{421} would find an explanation. With this in mind we can not exclude the fact that the two catalytic residues of Ldt_{fm} might behave similarly to His^{126} and Cys^{142} in Ldt_{Bs} despite the first impression that was limited to His^{421} aromatic ring at a single pH value.

7.2.2.3 Distribution of the three histidine acido-basic and tautomeric forms in Ldt_{fm} at pH 6.4

As detailed in section 4.2.4.2 on page 121, the fraction of each of the three possible forms of the imidazole ring can be determined at the pH of the NMR structural study, *i. e.* pH 6.4. The distribution

obtained for the four histidines of Ldt_{fm} are indicated in Table 7.1.

	His^{339}	**His^{353}**	**His^{421}**	**His^{440}**
$\delta_{obs-N\delta1}$	233.2	no fit	171.7	238.1
$\delta_{obs-N\epsilon2}$	179.2	no fit	no data	173.3
$\Delta\delta_{obs-C\delta2}$	0.58	no fit	jump	/
$f(N\epsilon2_{N\delta1})$	80.2 %	/	5.2 %	86.0 %
$f(N\epsilon2_{N\epsilon2})$	85.7 %	/	no data	92.9 %
$f(N\epsilon2_{C\delta2})$	83.5 %	/	no data	/
$f(N\epsilon2)$	83.1 ± 2.7 %	/	5.2 %	89.5 ± 3.4 %
K	0.20 ± 0.02	/	18.34	0.12 ± 0.05
pKa	6.85 ± 0.01	> 9	6.81 ± 0.07[1]	7.02 ± 0.03
$f_{acid-form}$	73.8 %	> 99 %	/	80.7 %
$f_{N\epsilon2-H-tautomer}$	21.8 %	< 1 %	/	17.3 %
$f_{N\delta1-H-tautomer}$	4.4 %	< 1 %	/	2.0 %

Table 7.1: Thermodynamic constants and fraction of the different protonation states and tautomeric forms of histidines in Ldt_{fm} at 25 °C and pH 6.4. δ_{obs-X} values are the experimental chemical shifts, where X is the Nδ1 or Nϵ2 nucleus and $\Delta\delta_{obs-C\delta2}$ represents the Cδ2 chemical shift difference at pH 6.4 and on the plateau at high pH; $f(N\epsilon2_X)$ stands for the fraction of the neutral basic Nϵ2-H tautomer, where X is the nucleus used as a probe. K and pKa values are obtained on the basis of the Nϵ2 average fraction $f(N\epsilon2)$ and on the titration curve fits, respectively. The final fractions of the three forms of the histidine imidazole ring in Ldt_{fm} are calculated using the equations described in section 4.2.4.2. Estimated errors on the final fractions are ± 5 %. [1] The pKa value given here may correspond either to a true pKa value for His^{421} or may reflect the impact of the acido-basic equilibrium of His^{440}. The answer to this question awaits a study on the H440A mutant.

The fraction of each acido-basic and tautomeric form of His^{421} escaped determination, as the determined pKa may well arise from indirect effects due to the changes in the protonation states of His^{440}. However, according to the chemical shift of the Nδ1 along the whole pH titration (172.6 ppm in acidic conditions and 171.7 ppm in basic conditions), Ldt_{fm} is likely to exist predominantly as the Nδ1-H-tautomer. Indeed, on the lone basis of the Nδ1-chemical shift at basic pH, the fraction of Nδ1-H-tautomer is estimated to be around 95 %. Yet, this fraction takes into account the hypothesis postulating that the protonation state of His^{421} remains the same in the whole pH range, which needs to be confirmed by the study of the H440A mutant.

Comparison with Ldt_{Bs} indicates that the fraction of charged acidic forms is larger in Ldt_{fm} (> 73 % for three of the four histidines). The study on the H440A mutant appears crucial to confidently answer the initially raised question on the protonation state of His^{421} and the existence of a hydrogen bond network with similar characteristics to what was described in Ldt_{Bs}. The determination of the protonation state for this histidine is essential for a good resolution of the active site in the NMR structure of the Ldt_{fm}

enzyme. Nevertheless preliminary structure calculations were performed in the mean time; they are described in the following section.

7.2.3 Determination of the NMR structure of Ldt_{fm}

Correlations in the 3D ^{13}C-methyl-selective-NOESY were manually assigned. The automatic peak-picking and NOE assignment of the ^{13}C- and ^{15}N-NOESY-HSQC spectra were performed using the iterative procedure of UNIO'10 (Herrmann *et al.*, 2002). ARIA2.3 was subsequently run with seven iterations calculating 100 structures, and a last iteration including 1000 structures. The 20 lowest energy structures were then refined with CNS using explicit solvent in a molecular dynamics simulation.

Three structure calculations were performed to take into account the three possible protonation states of His^{421}: the positively charged acid form, the neutral basic Nϵ2-H tautomer (even though it is the least probable state) and the neutral basic Nδ1-H tautomer. The protonation states of His^{440} correspond to the positively charged acidic form and the sulfur of Cys^{442} is protonated in all these calculations. No significant RMSD difference was observed between the three calculations, emphasizing the necessity to experimentally determine the His and Cys protonation states.

The superposition of the 20 lowest energy structures obtained after water refinement in the case of the Nδ1-H tautomer for His^{421} is shown in figure 7.11. This structure ensemble is well resolved, with a RMSD of 0.21 $\pm$ 0.03 and 0.47 $\pm$ 0.06 in the secondary structure elements for backbone and heavy atoms, respectively; and a RMSD of 0.47 $\pm$ 0.12 and 0.66 $\pm$ 0.07 on all residues of the protein for backbone and heavy atoms, respectively. However, this structure needs to be refined, as the NOE energy is still high (E_{NOE} = 2400 to compare with 657 in the case of Ldt_{Bs}). The determination of the protonation state of His^{421} will help to refine the structure of the active site and a search for additional long range NOE may help to fix the position of the long protein region containing the β5 and β6 strands that is acting like a "lid" and protects the active site from exposure to the solvent.

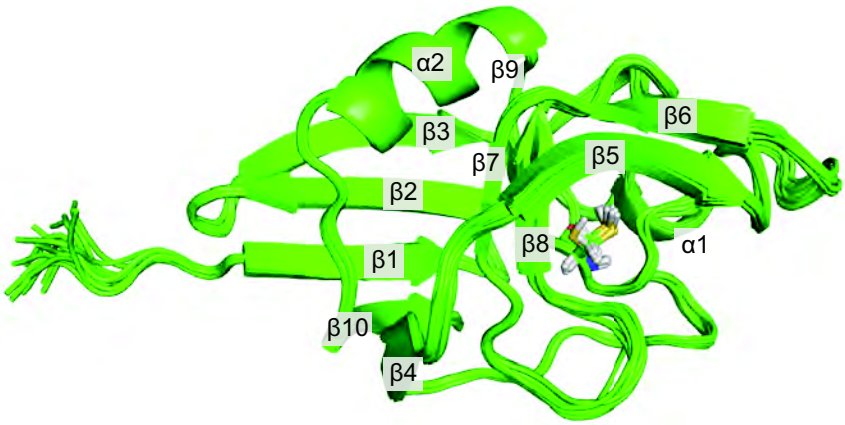


Figure 7.11: Superposition of the 20 lowest energy structures of the catalytic domain of Ldt_{fm}. Secondary structure elements are annotated, and the catalytic cysteine is shown as sticks. This calculation was run under the assumption that Cys^{442} occurs as a thiol and all histidines are in the protonated states, with the exception of the neutral basic Nδ1-H tautomer for His^{421}.

Comparison with the X-ray structure (Biarrotte-Sorin *et al.*, 2006) reveals that the general fold of the protein is preserved in solution in the absence of Zn^{2+}, with a global RMSD of 1.3 Å between the X-ray structure and the lowest energy NMR structure. Main differences arise in the orientation of the long region that covers the active site, including strands β5 and β6 and the loop between them. It might be necessary to record RDCs in order to fix the orientation of this lid, as only a few NOEs are detected between this particular region and the rest of the protein. This is a key issue as this region might have a role in the acylation mechanism.

7.3 Non-covalent interaction of Ldt_{fm} with the penam, cephem and carbapenem β-lactam families

Biochemical and kinetic studies led by the group of Michel Arthur have shown that Ldt_{fm} is able to catalyze the hydrolysis of the β-lactam ring in a variety of classes, but with different efficiencies. Acylenzymes are formed with meropenem (carbapenem family), faropenem (penem family), cefotaxime (cephem family), and ampicillin (penam family). Nevertheless the antibiotic skeleton is degraded in some cases (faropenem and ampicillin) and the enzyme turnover greatly differ: 800, 1000, 2500, 105 min^{-1} for meropenem, faropenem, cefotaxime and ampicillin respectively. The specificity for carbapenem needs to be investigated. NMR provides a convenient tool to screen for specificity at the stage of the formation of the non-covalent Michaelis-Menten complex. A Ldt_{fm} catalytic domain was thus prepared by the group of Michel Arthur with the C442A mutant to screen for non-covalent interaction with penam, cephem and carbapenem. Four equivalent C442A Ldt_{fm} samples were thus prepared at the concentration of 150 μM in 100 mM sodium phosphate buffer containing 300 mM NaCl at pH 6.4.

The C442A substitution did not affect the conformation of the protein since ^{1}H-^{15}N-HSQC spectra of C442A-Ldt_{fm} and *wt*-Ldt_{fm} are highly superimposable, and only minor chemical shifts are observed for residues in close proximity to residue 442 (data not shown).

A twenty-minutes long ^{1}H-^{15}N-HSQC spectrum was collected at 25 °C on a 600 MHz proton frequency Agilent spectrometer equipped with a cryogenic probe after each successive addition of antibiotic, first through addition of small volumes from stock solution, and then by addition of antibiotic powder in the NMR tube in order to prevent the sample from significant dilution. Four antibiotics from three different β-lactam families were tested: ampicillin as a representative of the penam family, ceftriaxone as a representative of the cephem family, and imipenem and ertapenem to probe for the carbapenem family.

The last point of the four titrations was dictated by the limit of solubility of the different antibiotics: ∼ 300 mM for ampicillin (2,058 equivalents of antibiotic compared to the enzyme), ∼ 150 mM for ceftriaxone (1,025 equivalents), ∼ 50 mM for imipenem (328 equivalents) and ∼ 300 mM for ertapenem (2,016 equivalents). A superposition of the ^{1}H-^{15}N-HSQC spectra collected along the 15 titration points of Ldt_{fm}-C442A by ertapenem is shown in figure 7.12. The interaction of Ldt_{fm}-C442A with the antibiotics occurs in the fast exchange regime, except in the case of ceftriaxone where significant intensity changes have been noticed for some of the resonances arising from residues located at the vicinity of Ala^{442}. The interaction study by NMR is limited in the case of imipenem due to its low solubility at pH 6.4 and 25 °C. None of the chemical shift perturbation was significant enough to lead to a structural or

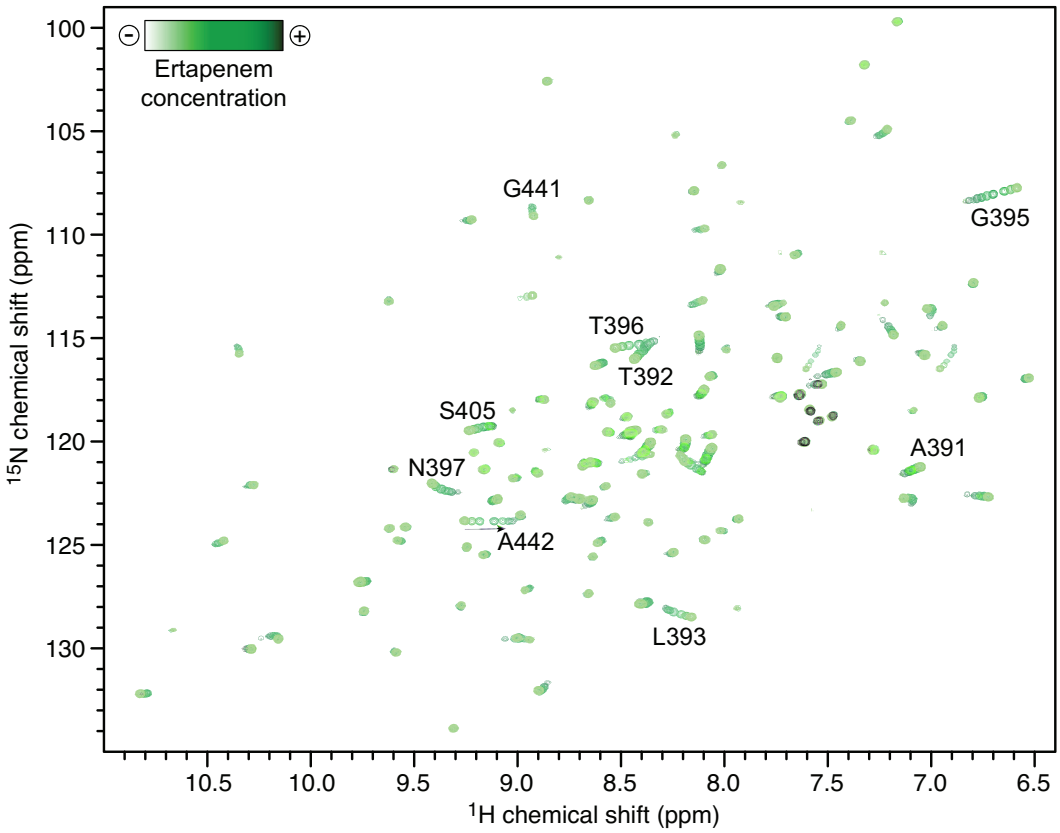


Figure 7.12: ^{1}H-^{15}N-HSQC spectra of Ldt_{fm} along the titration with ertapenem in phosphate buffer at pH 6.4 and 25 °C. The reference spectrum was recorded in the absence of ertapenem (in clear green), and successive ertapenem additions were performed between collected datasets until an antibiotic-to-protein molecular ratio of 2,016 was reached (in dark green). Protein concentration was 150 µM and large quantities of antibiotics were added in the solid state to avoid protein dilution along the titration. The residues that are significantly moving along the titration are labeled on the spectrum.

quantitative information. The results for this antibiotic are thus omitted in the rest of the discussion.

For each residue, the chemical shift perturbation at a given titration point is calculated according to equation 3.9 on page 101. To extract structural information on the interaction site in a comparative manner for the three antibiotics, a common titration point was chosen, corresponding to a ligand-to-protein ratio of 1,000 (figure 7.13a), and medium-to-strong CSPs were reported on the apoenzyme NMR structure (figure 7.13c). The residues showing an intermediate exchange regime in the case of the ceftriaxone titration are also clearly identified on the structure.

Three regions of the protein are consistently affected by all three antibiotics (figure 7.13a). The first region stands from residues 391 to 397 and residues 401 to 405, which are located at the beginning of the lid. The second region ranges from residues 438 to 442, which corresponds to the loop containing the catalytic cysteine. Residues 414 and 415 (located between strands β7 and β8) are also systematically affected by the antibiotics. These sites could be seen as primary specific interaction sites for all antibiotics. Then, the chemical shift mapping presented in figure 7.13c clearly emphasizes differences in the antibiotics interaction modes. Interaction with ampicillin for example exclusively affects the other side of Ldt_{fm}-C442A, with CSPs of medium amplitude in strands β1, β10 and β4. For both ampicillin and ertapenem, residues 422 to 426 encountered medium CSPs, and His^{421} encountered strong CSPs,

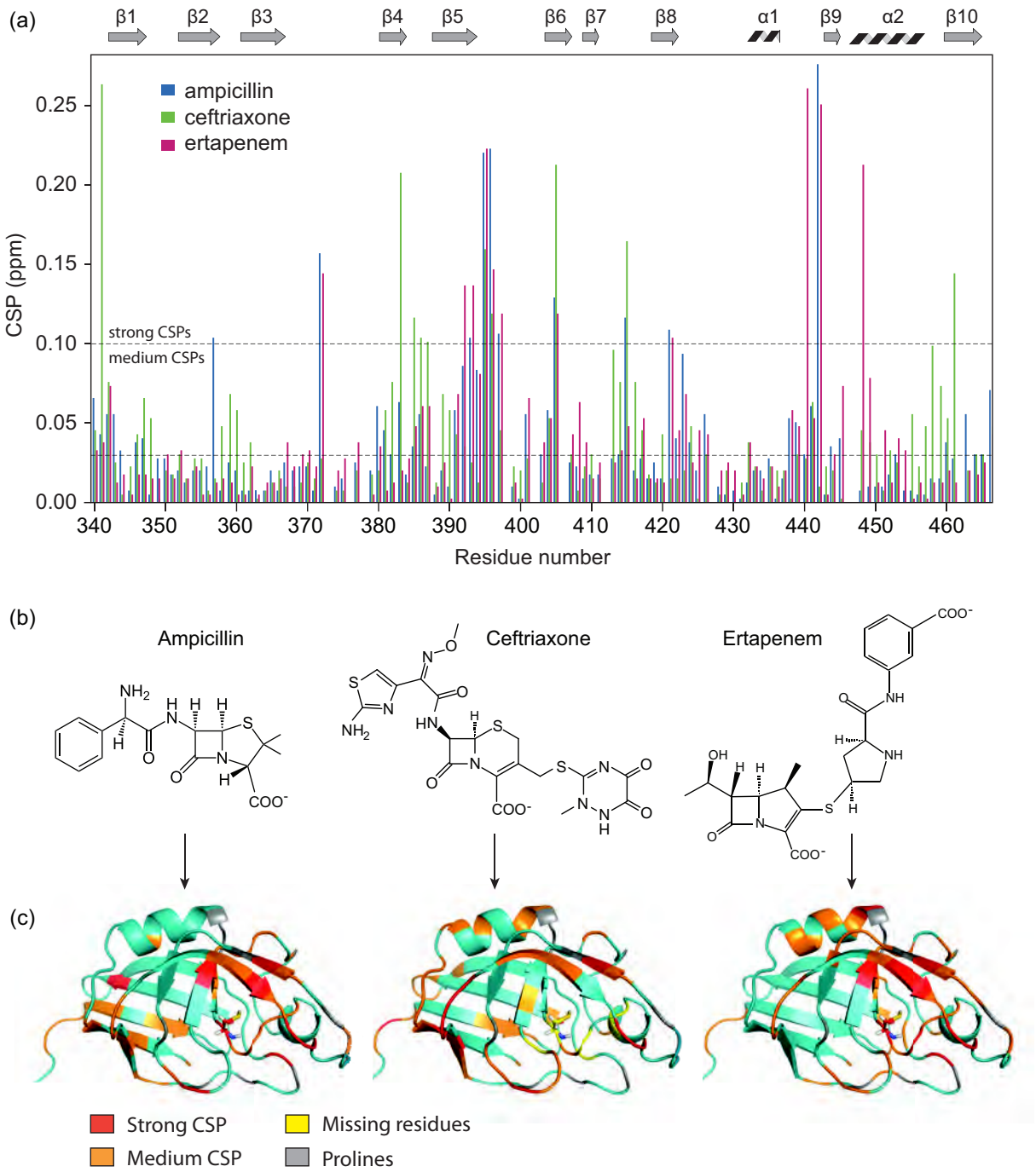


Figure 7.13: Chemical shift perturbations induced by the non-covalent interaction between Ldt_{fm}-C442A and antibiotics from three different β-lactam families: ampicillin (penam family), ceftriaxone (cephem family), and ertapenem (carbapenem family). a) CSPs induced by the addition of 1,000 equivalents of ampicillin (in blue), ceftriaxone (in green), and ertapenem (in pink). **b)** Chemical formula of the three β-lactam antibiotics. **c)** CSPs induced by the addition of 1,000 equivalents of each antibiotic reported on the Ldt_{fm} apoenzyme NMR structure. Residues with medium (0.03 ppm < CSP < 0.1 ppm) and strong (CSP > 0.1 ppm) CSPs are colored in orange and red, respectively. For ceftriaxone, some residues become less intense along the titration, until disappearance in some cases, suggesting an intermediate exchange regime. Corresponding missing residues at 1,000 equivalents are colored in yellow. Prolines are colored in grey, and the catalytic cysteine is shown as sticks.

while this region does not seem affected by ceftriaxone. With ertapenem, the helix α2 is more affected than with the two other antibiotics. These modulations may arise from secondary effects that occur upon binding and may be related to slight conformational rearrangement of the protein. The antibiotic affinity is too weak to actually determine a structure of the non-covalent complex. The present qualitative results nevertheless suggest that each of the antibiotic specifically binds to the protein active site, and that the steric hindrance on the β-lactam ring induces remote mild conformational rearrangement. In this hypothesis, the R^2 substituent (see figure 2.15 for the nomenclature) would point towards the back side of the protein (strands β1, β4 and β10) while a R^1 substituent would induce conformational perturbation in the direction of helix 2. Interestingly the ceftriaxone which contains substituents on both sides of the β-lactam ring presents characteristics of the two previous β-lactam families. Alternatively the occurrence of a second binding site with weaker affinity could be proposed in the case of ampicillin, and to a lower extent in the case of ceftriaxone.

To try to differentiate these two working hypothesis, titration curves were thus analyzed. In each case, residues that showed a CSP greater than 0.03 ppm at the end of the titration were selected. The evolution of the CSP as a function of ligand-to-protein ration ($[L_0]/[P_0]$) of all selected residues was fitted simultaneously with a single dissociation constant according to equation 4.9 (on page 111). In the case of ceftriaxone, the fit included 24 residues (340, 372, 381-383, 385-387, 395-397, 401, 405, 415, 416, 420, 439-441, 457-461) and yielded to a K_D of 43 ± 1 mM. In the case of ertapenem, the fit included 20 residues (350, 367, 372, 391-395, 397, 401, 403-405, 408, 409, 438, 439, 441, 442, 453) and yielded to a K_D of 50 ± 2 mM. Examples of this quantitative analysis are presented in figure 7.14. A similar fitting procedure did not offer a conclusive dissociation constant for the ampicillin case in the frame of this 1:1 binding model. Analysis of individual titration curves suggest that a 1:2 binding model might be more adequate. This analysis is still in progress.

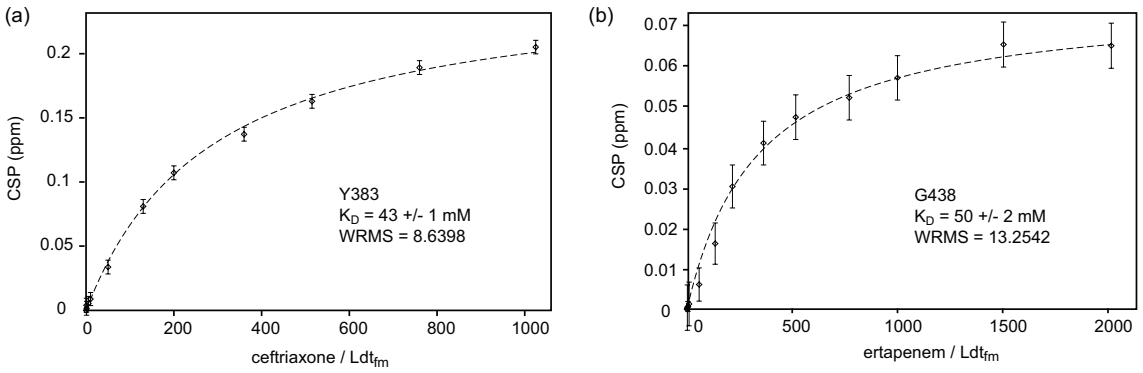


Figure 7.14: Quantitative CSP analysis as a function of ligand-to-protein ratio and determination of the dissociation constant of the non-covalent complex between Ldt_{fm}-C442A and β-lactam antibiotics. a) Experimental data points (represented as diamonds with their respective error bars) and the result of the fitting procedure (dotted line) led simultaneously on 24 residues in the case of Tyr^{383} for the Ldt_{fm}-ceftriaxone non-covalent complex. **b)** Experimental data and the result of the fitting procedure led on 20 residues in the case of Gly^{438} for the Ldt_{fm}-ertapenem non-covalent complex.

The dissociation constants extracted from the non-covalent interaction of cephems and carbapenems with C442A-Ldt_{fm} are essentially identical (approximately 50 mM) and indicate that the affinity is slightly stronger than the non-covalent interaction of penams and carbapenems with C142A-Ldt_{Bs} ($K_D \geq 100$ mM). Yet, the interaction is still considered as a low-affinity interaction, with K_D values in the milli-molar range. In addition to the K_D values, the localization of the three antibiotics in the vicinity

of the active site suggests that, similarly to Ldt_{Bs}, the non-covalent interaction is not the determinant step in the acylation pathway of Ldt_{fm}.

7.4 Impact of acylation by carbapenems on the structure and dynamics of Ldt_{fm}

7.4.1 Assignment of NMR resonances in the Ldt_{fm}-ertapenem acylenzyme

^{13}C-^{15}N-labeled Ldt_{fm} in 100 mM sodium phosphate buffer containing 300 mM NaCl at pH 6.4 was incubated with 2.5 equivalents of ertapenem, yielding to the acylenzyme within minutes. Due to limited chemical shift changes between the apo- and acylenzymes (figure 7.15), 3D HNCACB and HNCO experiments, collected on an Agilent spectrometer operating at 600 MHz proton frequency equipped with a cryoprobe, were sufficient to perform unambiguous backbone assignments.

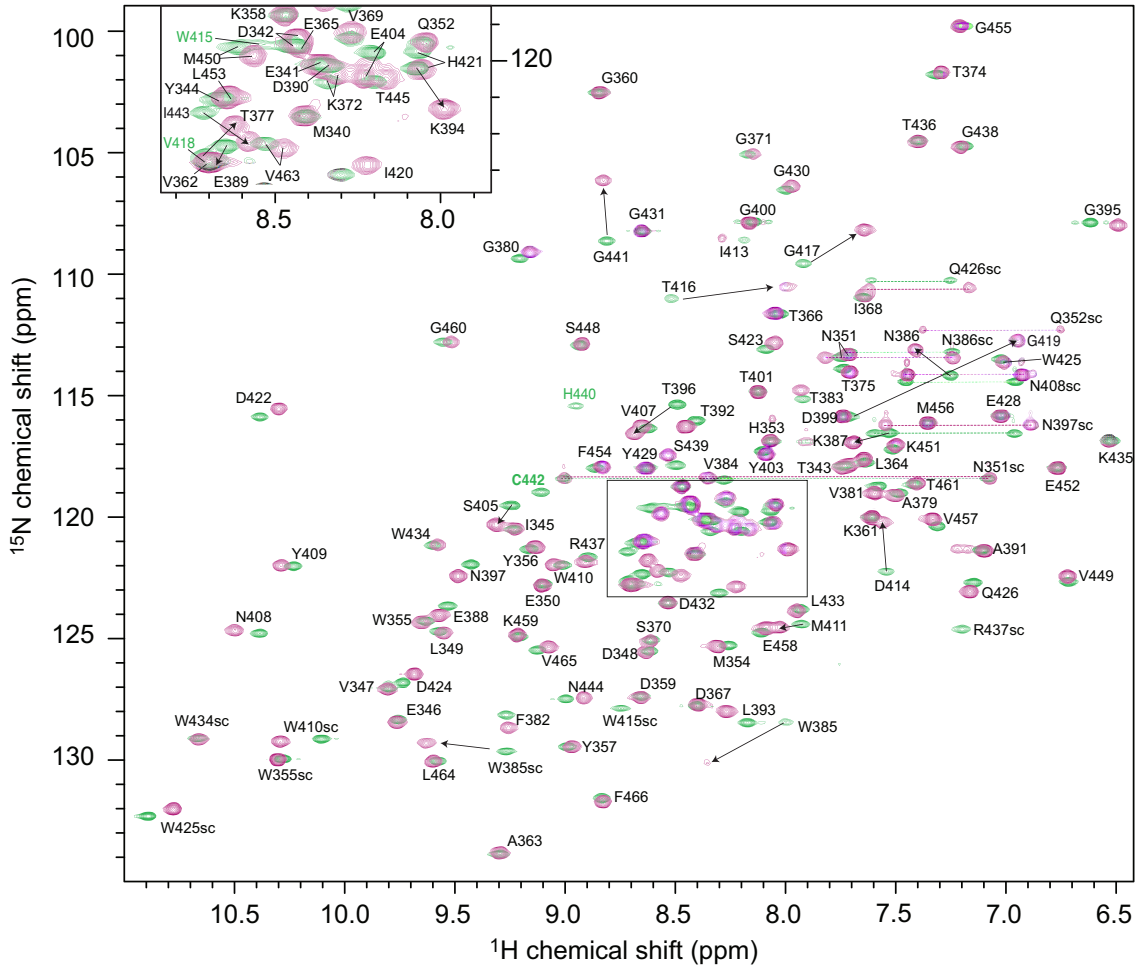


Figure 7.15: Superposition of the ^{1}H-^{15}N-HSQC spectra of the apo (in green) and ertapenem-acylated (in purple) Ldt_{fm} enzyme in 100 mM sodium phosphate, 300 mM NaCl buffer at pH 6.4 and 25 °C.

For the side-chains assignment of ertapenem-bound Ldt_{fm}, a HCCH-TOCSY and a ^{15}N-NOESY-HSQC (τ_m = 150 ms) were recorded on an Agilent Direct Drive 600 MHz spectrometer equipped with a cryogenic probe. ^{13}C-NOESY-HSQCs centered on aliphatic and on aromatic nuclei were recorded

with respective mixing times of 130 ms and 155 ms. These ^{13}C-HSQC-NOESYs were recorded on the national TGIR (Très Grandes Infrastructures de Recherches) platform at Gif sur Yvettes, on a Bruker Avance spectrometer operating at 950 MHz equipped with a TCI cryogenic probe.

The overall backbone assignments on Ldt_{fm} acylated by ertapenem were completed to 95.3 % for the non-proline $^{1}H_N$-^{15}N resonances (with the exception of the N-terminal residue), 93.8 % for the $^{13}C\alpha$, 96.5 % for the $^{13}C\beta$ and 81.2 % for the ^{13}C' resonances. In total, 93.8 % of the ^{1}H side-chain resonances were assigned in the acylenzyme. The chemical shifts are in the process of being deposited in the BMRB database.

7.4.2 Structural rearrangement induced by ertapenem

The CSPs induced by the acylation of Ldt_{fm} by ertapenem are reported in figure 7.16.

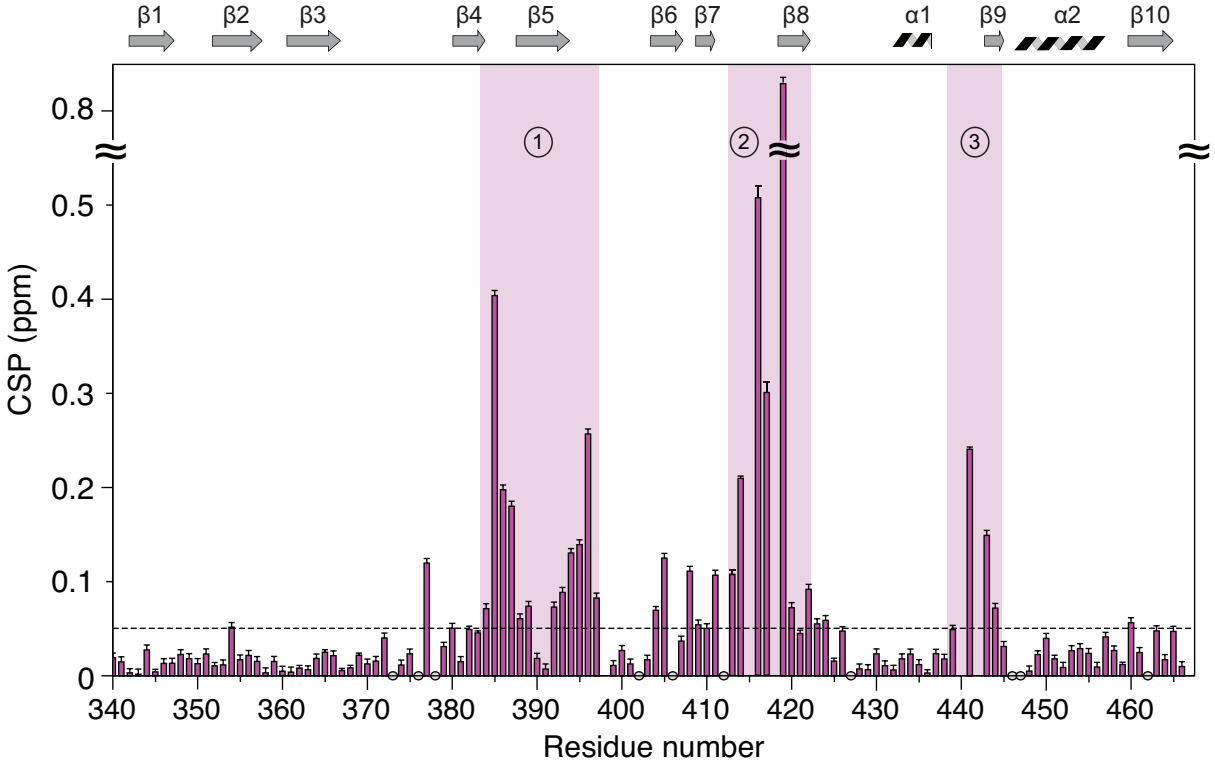


Figure 7.16: Chemical shift perturbations observed upon acylation of Ldt_{fm} by ertapenem antibiotic. Amide resonances that remain undetected in the ^{1}H-^{15}N-HSQC spectrum of the acylenzyme account for the missing data at position 415, 418, 440 and 442, except for residue 398 which was already undetected in the apoenzyme ^{1}H-^{15}N-HSQC spectrum and which is located in the middle of the lid. Three regions are particularly affected, from residues 384 to 389 and 392-397 (region 1), from residues 411 to 423 (region 2) and from residues 439 to 443 (region 3). Prolines are represented as grey circles.

Figure 7.16 shows that the acylation of Ldt_{fm} by ertapenem induces significant structural rearrangement, with CSPs up to 0.8 ppm and with approximately the quarter of the residues with strong CSP values (above 0.05 ppm). Residues particularly affected by the covalent binding, that show strong chemical shift changes or increased dynamics (missing resonances), are located within 3 regions. Region 1, from residues 384 to 397, includes the β5 strand from the lid. Region 2, from residues 413 to 423, includes the loop following the β7 strand, and the strand β8 that contains the putative catalytic His^{421}. This

portion faces the loop containing the catalytic Cys^{442}. Finally, the latter loop constitutes region 3, from residues 439 to 443.

CSP values are similar to those previously obtained upon acylation of Ldt_{Bs} by imipenem, which suggests that the structural rearrangements of Ldt_{fm} enzyme will share common characteristics and will be reduced to the three aforementioned regions. Regions 2 and 3 in Ldt_{fm} clearly overlap with the corresponding regions in Ldt_{Bs} as suggested from protein sequence alignments and from the analysis of the NMR or X-ray structures (see figure 7.17). Region 2 in Ldt_{fm} is nevertheless more extended (13 residues instead of 8), while region 3 is more restricted (5 residues instead of 14). In contrast region 1 differ in both acylenzymes. In Ldt_{Bs}, this region was the equivalent of residues 370 to 380 in Ldt_{fm}, a loop that is found below the active site. On the contrary, region 1 in Ldt_{fm} corresponds to a loop located above the active site. This difference may reflect a different positioning of the antibiotics in the acylenzyme and need to be further investigated.

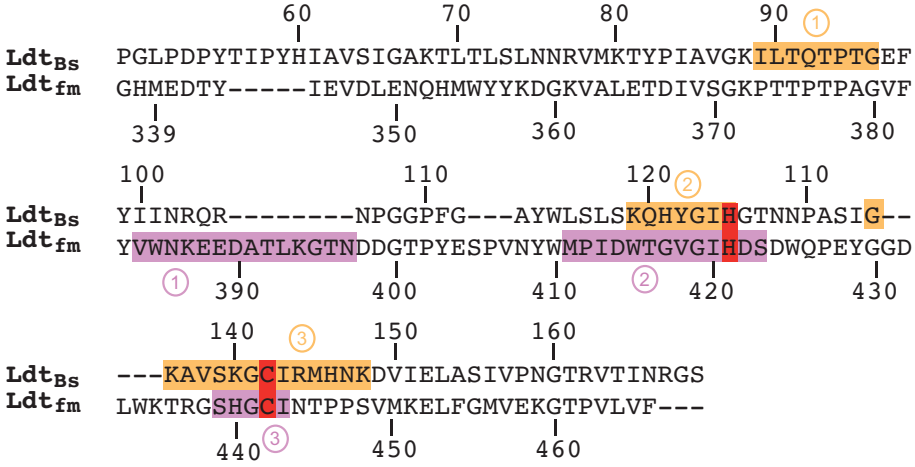


Figure 7.17: Sequence alignment of the catalytic domains of Ldt_{Bs} and Ldt_{fm} showing the regions affected by carbapenem acylation. Ldt_{Bs} and Ldt_{fm} enzymes contain three particularly affected regions, highlighted in orange and in purple, respectively. The catalytic histidine and cysteine residues are highlighted in red.

Another significant difference between the two acylenzymes stands in the difference of dynamics. In the Ldt_{Bs} acylenzyme, 22 NH correlations were missing in the ^{1}H-^{15}N-HSQC spectrum upon acylation by imipenem, due to conformational exchange in the ms time-scale present in all three perturbed regions. This number must be put in parallel to the only 4 missing NH correlations in the Ldt_{fm} acylenzyme that are distributed only in regions 2 and 3. Does this difference stand in the time-scale or the extent of the motion? To address this question, relaxation measurements were performed on both the apo- and the acyl-Ldt_{fm}-enzyme. They are detailed in the following section.

7.4.3 Comparison of Ldt_{fm} dynamics before and after acylation by ertapenem

The NMR study of Ldt_{fm} dynamics was performed on ^{13}C-^{15}N-samples of both 930 µM apo- and 700 µM imipenem-acylated Ldt_{fm} 100 mM sodium phosphate buffer containing 300 mM NaCl at pH 6.4. R_1, R_2 and heteronuclear NOE experiments were collected at 25 °C on an Agilent Direct Drive

spectrometer operating at a proton frequency of 600 MHz and with a non-cryogenic probe to avoid heating probe issues. Data processing and analysis were performed using NMRPipe (Delaglio *et al.*, 1995) and in-house modified scripts..

7.4.3.1 Determination of ^{15}N-relaxation parameters

To determine the two relaxation rates R_1 and R_2, a series of 2D ^{1}H-^{15}N-correlation spectra were recorded with 15 different relaxation delays t (see section 5.2.1) in order to get a good sampling of the exponential-decay curves, yielding to an accurate curve fitting. ^{15}N-R_1 values were measured with 0, 20, 50, 100, 200, 300, 400, 500, 700, 900, 1200, 1500 and 1800 ms relaxation delays, and ^{15}N-R_2 values were measured with 10, 30, 50, 70, 90, 110, 130, 150, 170, 210 and 250 ms. To estimate experimental errors, two spectra were recorded with a redundant relaxation delay: 200 ms and 700 ms for R_1, and 70 ms and 210 ms for R_2.

The intensity of each amide correlation followed a mono-exponential decay as the relaxation delay was increased (see section 5.1). Results of the fits and estimate of errors obtained from a Monte-Carlo simulation are reported in figure 7.18.

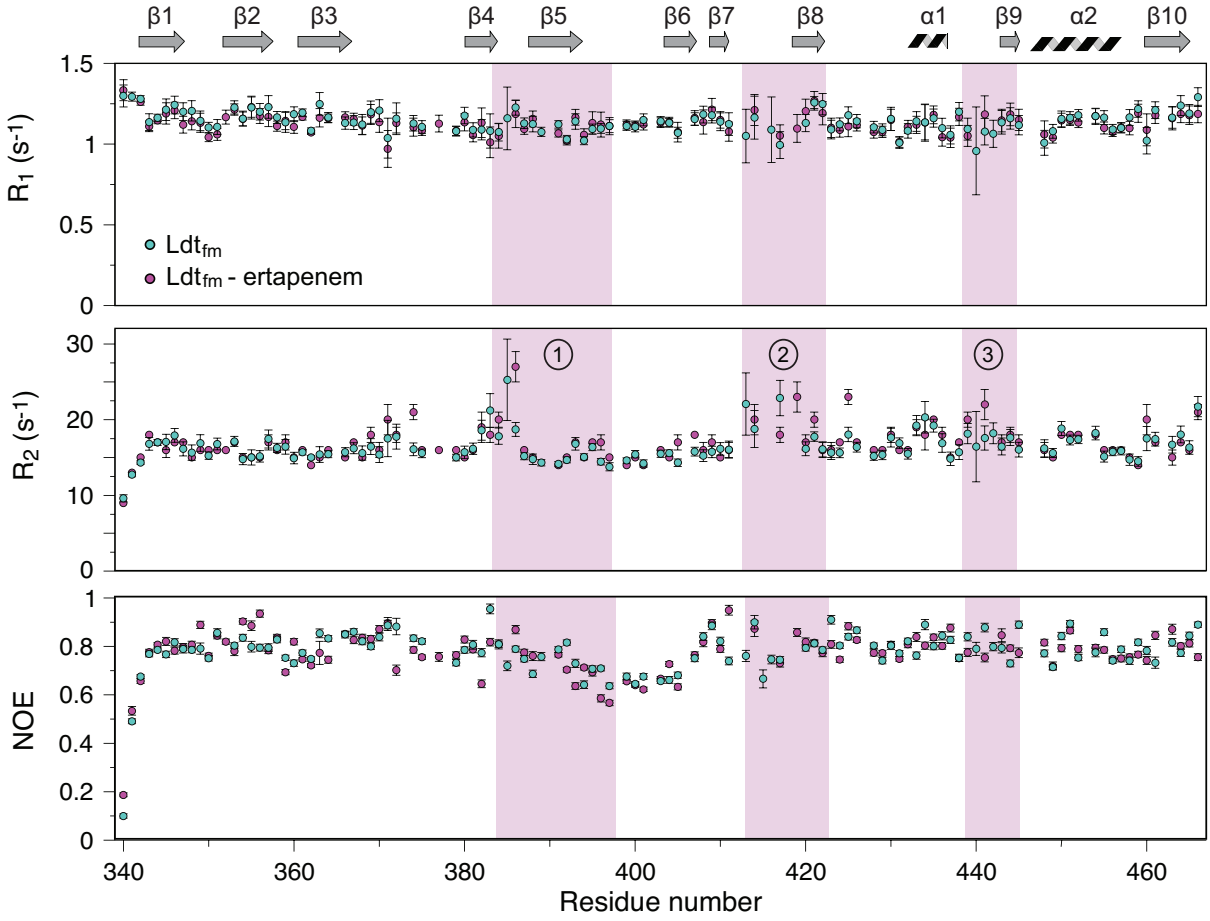


Figure 7.18: ^{15}N relaxation parameters R_1, R_2 and NOE of Ldt_{fm} apoenzyme in blue teal and the ertapenem-acylated enzyme in magenta. The three affected regions by structural rearrangement are shown.

For ^{1}H-^{15}N-heteronuclear NOE measurements, a reference and an amide-saturation spectrum were collected with a 2 s relaxation delay and a 3 s on- and off-resonance period, respectively. The NOE values were calculated according to equation 5.12 and are depicted for each residue in figure 7.18.

7.4.3.2 Interpretation of ^{15}N-relaxation parameters

Relaxation data are missing for proline residues, in position 373, 376, 378, 402, 406, 412, 427, 446, 447 and 462, as they do not contain amide protons. The N-terminal residues G338 and H339 as well as residue D398 are not observable in ^{1}H-^{15}N-HSQC spectra. In the apoenzyme, the residues Q352, E365, L453, T377, V418 and G419 were involved in overlaps and were thus removed from the analysis. In the ertapenem-acylated enzyme the residues E341, E365, E389, D390 and L453 were involved in overlaps and excluded from the dynamics analysis as well as residues W415, V418, H440 and C442 that remained undetected in the ^{1}H-^{15}N-HSQC spectrum. Furthermore, the intensity of the signals arising from Ile^{413} and Thr^{416} were also too weak to yield reliable interpretation of the relaxation data.

The average value for the relaxation rate constant R_1 is homogeneous along the sequence as shown in figure 7.18, with an average value of 1.14 $\pm$ 0.058 s^{-1} for both forms of Ldt_{fm}. Only few residues present an R_1 value outside from more than one average standard deviation (0.07 s^{-1}). These residues include the N- and C-terminal ends of the protein and some isolated residues (363, 386, 394, 422 for the apoenzyme; 371 for the acylenzyme; and 353, 355, 392, 421, 431 in both forms).

Heteronuclear NOE data reveal a globally rigid protein, with an average value of 0.778 $\pm$ 0.015 for the apoenzyme, and of 0.775 $\pm$ 0.015 for ertapenem-acylated Ldt_{fm}. Fast time-scale (ps to ns) dynamics ($NOE \leq 0.72$) is observed only at the N-terminal part of the protein (residues 340 to 342), and in the lid (residues 394 to 405) in both forms of the enzyme.

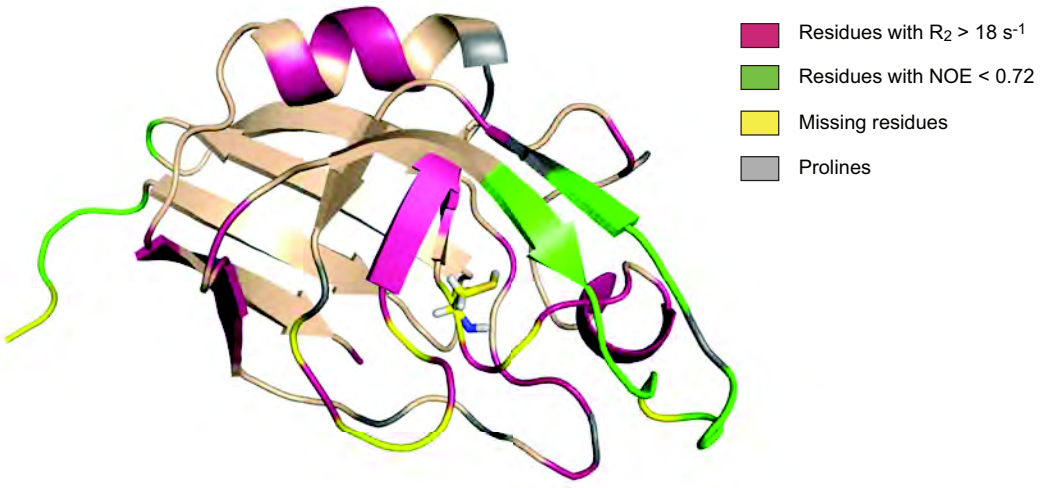


Figure 7.19: Regions exhibiting dynamics at different time-scales on ertapenem-acylated Ldt_{fm}. Fast (ps to ns) time-scale dynamics is identified through a heteronuclear NOE value lower than 0.72 (corresponding residues are colored in green), and slow (micros to ms) time-scale dynamics is identified through R_2 values greater than 18 s^{-1} (corresponding residues are colored in pink on the structure). Missing residues (in yellow) are detailed in the main text. The structure used for this figure is the NMR structure of the Ldt_{fm} apoenzyme.

Finally, R_2 relaxation rates reveal significant exchange contributions in the apoenzyme of Ldt_{fm}, in regions 1 and 2 located at the vicinity of the catalytic cysteine. Both the time-scale and the extension of these motions seem to slightly shift upon acylation. Figure 7.19 sketches on the structure of the apoenzyme the regions affected either by fast (in green, according to NOE values) or slow dynamics (in pink, according to high R_2 values).

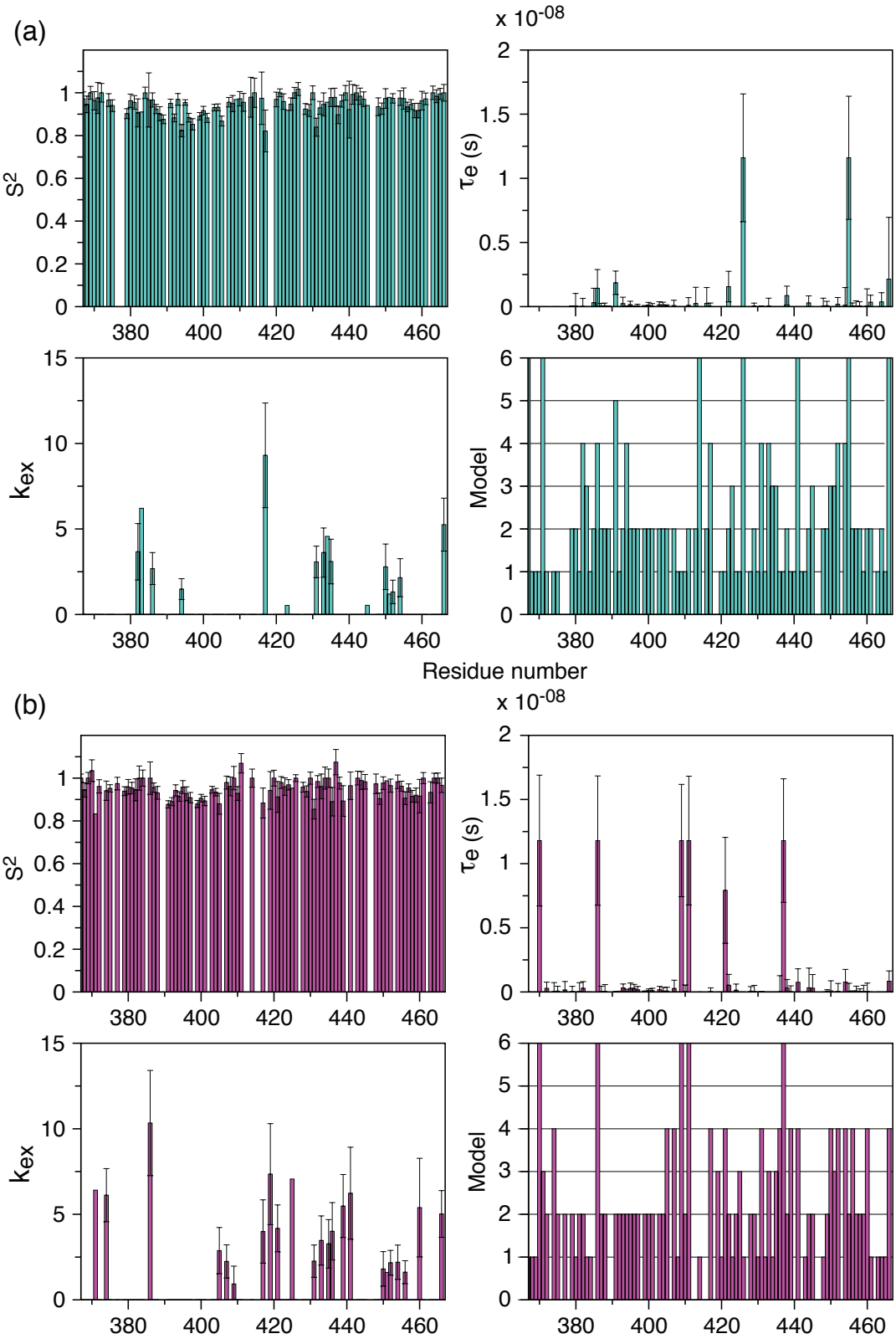


Figure 7.20: Model-free analysis of the internal mobility of Ldt_{fm} using TENSOR2 for a) the apoenzyme and b) ertapenem-acylated enzyme.

Quantitative analysis of the present relaxation data was performed using TENSOR2 (Dosset *et al.*, 2000) for both the apo- and the acyl-enzyme (figure 7.20). On the contrary to Ldt_{Bs}, that contains an elongated LysM domain, an isotropic tensor was used to interpret the relaxation data on the catalytic domain of Ldt_{fm}. The value of the correlation time τ_c deduced from this analysis was 11.6 ns for the apoenzyme and 11.8 ns for the ertapenem-acylated enzyme. These values are unusually high for a molecule of this shape and size. A calculation of τ_c using HYDRONMR (García de la Torre *et al.*, 2000) yielded to a correlation time of 8.5 ns, which is more in accordance with the expected value. The difference of a factor 1.35 can be explained by the presence of glycerol in the Ldt_{fm} sample, that affects the viscosity of the solution and was proven by the collection of a simple 1D ^{1}H NMR spectrum. The viscosity of water is 1.005 cP (centiPoise) at 20 °C, while it increases to 1.31 and 1.76 cP in the presence of 10 and 20 % of glycerol, respectively. It decreases when temperature increases (Segur and Oberstar, 1951). In isotropic conditions, equation 5.6 (page 127) shows that the correlation time is directly proportional to the viscosity. We can therefore estimate the quantity of glycerol to be between 10 and 15 % from the relaxation data.

In addition to the correlation time, TENSOR2 provides us with a comparative analysis of the motions time-scale and amplitudes between the two forms of the Ldt_{fm} enzyme. Fast ps-to-ns time scale motions have similar amplitudes in both states and are extremely limited, as shown by the S^2 parameter. Slow (micros-to-ms) motions preexist in all three regions 1-3 in the apo-enzyme and tend to diminish in region 1 while it extends around regions 2 and 3 in the acylated enzyme (see the k_{ex} graphs in a and b). These data need to be confirmed by relaxation-dispersion measurements in order to further characterize the time-scale of the exchange contribution detected in the R_2 profile. Nanoseconds motions described in the τ_e graphs are isolated and will not be described further, as they are correlated to a model 6 that indicates a poor fitting quality for these residues.

These results indicate that the behavior of Ldt_{fm} with respect to dynamics drastically differs from Ldt_{Bs}. First, the Ldt_{Bs} apoenzyme was shown to be globally rigid, leading to reject the preexistence of an excited state. Second, the impact of acylation was much more extended in the case of Ldt_{Bs}, with more than 20 residues affected by a large intensity decrease and significant ms time-scale motion. On the contrary, in Ldt_{fm}, both apo- and acyl-enzymes show similar exchange contributions, although these contributions are slightly more significant in the acylenzyme. Pursuing relaxation-dispersion experiments is crucial at this stage to examine the relative motional time-scales. Measurement at different fields in addition to what was done on Ldt_{Bs} would potentially offer a determination of the population of the preexisting excited state in the apoenzyme. A comparison with the acylated enzyme would also bring additional perspectives.

7.5 Conclusions on Ldt_{fm} and comparison to Ldt_{Bs}

Comparison between Ldt_{fm} and Ldt_{Bs} enzymes reveals both similarities and differences. First, in the apoenzyme, it is possible that the catalytic triad is conserved, with the Cys^{442}-His^{421}-Asp^{422} in the same configuration than Cys^{142}-His^{126}-Gly^{127}, with the putative catalytic histidine protonated on the Nδ1. Nevertheless the exact protonation states of His^{421} and Cys^{442} await to be confirmed through a pH-titration study on the H440A mutant. Structure calculations with the Cys^{442} thiolate need to be performed and compared to the corresponding calculations with a thiol instead. In spite of a structural

analogy, differences in the dynamical behavior of the two apoenzymes were detected: while the Ldt_{Bs} apoenzyme was globally rigid and did not encounter any conformational exchange, relaxation measurements performed on Ldt_{fm} showed the presence of a preexisting conformational exchange contribution at the micro- to milli-second time-scale, in two regions located at the vicinity of the catalytic cysteine. This suggests that Ldt_{fm} could bind the β-lactam antibiotics through a selected-fit model rather than the close to induced-fit model that was proposed for Ldt_{Bs}. Then, the study of the non-covalent complexes demonstrated that in both enzymes the affinity of LDts towards β-lactams is rather weak ($\sim$ 100 mM for penams and carbapenems with Ldt_{Bs}, and $\sim$ 50 mM for cephems and carbapenems with Ldt_{fm}), but specific to the active site containing the catalytic cysteine. Furthermore the LDt specificity for carbapenem does not seem to occur at the stage of the formation of the Michaelis-Menten non-covalent complex. In the case of Ldt_{fm}, the binding of different β-lactams may nevertheless induce slight conformational changes remote from the binding site, which may then impact on the kinetic of the covalent bond formation between the enzyme sulfur and the β-lactam carbonyl group. Finally, the acylation by a carbapenem of both LDts causes similar structural rearrangements, as emphasized by the proximity of CSP profiles both in terms of localization of the strong CSPs and CSP absolute value. However, dynamics seem to affect Ldt_{fm}-ertapenem acylenzyme to a far less extent in the covered time-scale regime, as only a few residues disappear from the ^{1}H-^{15}N-HSQC spectrum (disappearance of 4 residues only against 22 in the case of the Ldt_{Bs}-imipenem acylenzyme). The difference in R_2 relaxation rates between Ldt_{fm} apo- and acyl-enzyme is also quite modest (although significant). Relaxation dispersion must be performed in order to quantitavely characterize the time-scale of the conformational exchange contributions detected in both Ldt_{fm} apo- and acyl-enzyme. We can take advantage of the minor Ldt_{fm} dynamics in comparison to Ldt_{Bs}, as it opens the way to obtain the first high-resolution structure of an acylenzyme in the case of Ldt_{fm}. A ^{13}C-filtered NOESY experiment centered on aliphatic protons (see section 3.2.2.2) was recorded in D_2O on the Ldt_{fm}-ertapenem acylenzyme, on an Agilent spectrometer operating at 800 MHz proton frequency, with a mixing time of 180 ms. The assignment of the unlabeled ertapenem molecule bound to Ldt_{fm} could be partially performed using the spectrum that was filtering contacts between protons attached to natural abundance carbon (^{12}C) and nitrogen (^{14}N) nuclei. Methyls H9m*, H10m* and protons H1, H6, H7 and H8 were detected in this spectrum and assigned (see figure 7.21), while aromatic protons H22, H23, H24 and H26 were detected but could not yet be assigned.

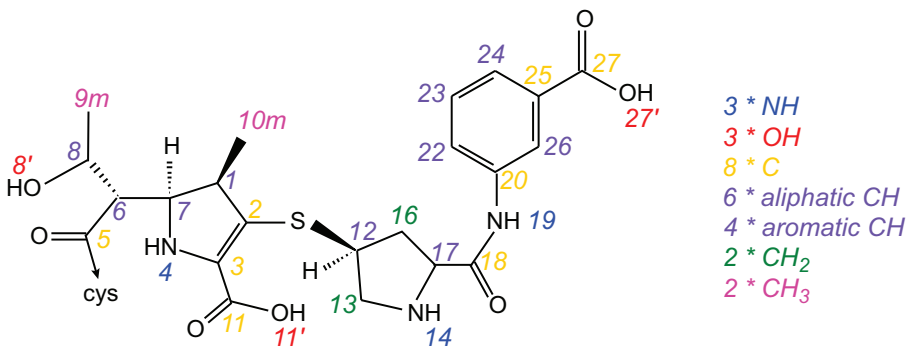


Figure 7.21: Representation of ertapenem with the opened β-lactam ring occurring upon covalent binding to the sulfur atom of the catalytic cysteine of Ldt_{fm}. Numbering used for the NMR assignment is provided. Assignment of protons attached to carbons 1, 6, 7, 9 and 10 was performed using the ^{13}C-filtered-NOESY experiment through the edition of the sub-sub-spectrum containing NOE information between protons attached to ^{12}C carbon atoms.

Then, intermolecular NOE correlations between the labeled protein and the unlabeled ertapenem antibiotic were identified in the "half-filtered" experiment. If these correlations need to be completed, particularly intense NOE correlations were detected between the two methyl groups of Leu^{393} in Ldt_{fm} on one side, and the H9m* methyl groups and H8 proton in the ertapenem in the other side. Leu^{393} is located in the β5 strand that covers the active site. These preliminary contacts are promising in the perspective of determining the structure of the acylenzyme, with a precise localization and orientation of the ertapenem within the active site. A modified-cysteine containing the opened ertapenem β-lactam ring was created within ARIA, in the same way as for imipenem in the case of Ldt_{Bs} (see modified files in Annexe 2). A preliminary structure calculation was performed on the Ldt_{fm}-ertapenem acylenzyme, and the PDB file obtained was used to generate the expected contacts between Ldt_{fm} and the ertapenem using the program Ligplot (Wallace *et al.*, 1995) (figure 7.22).

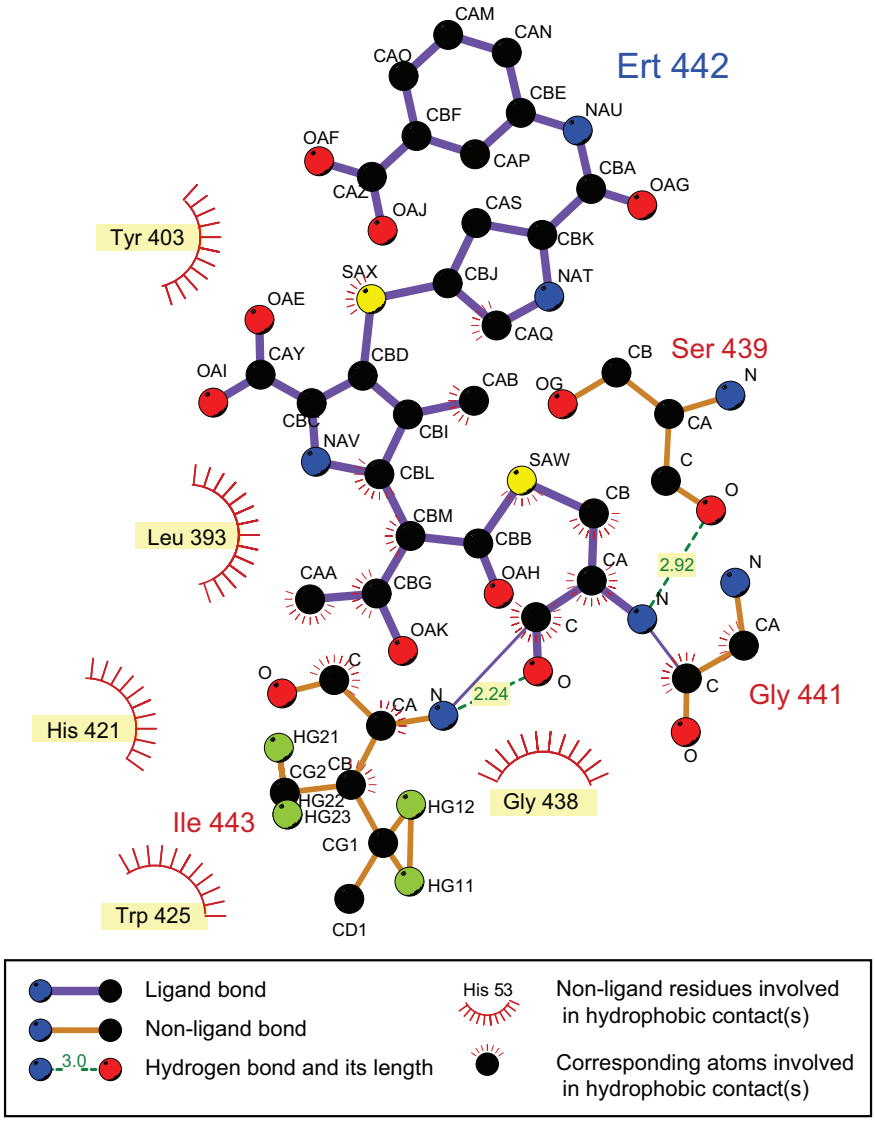


Figure 7.22: Potential contacts between the ertapenem antibiotic and Ldt_{fm} based on isotopically filtered NOE data and the correlated acylenzyme structure of lower energy as determined by ARIA, using LigPlot (Wallace *et al.*, 1995). Ert442 represents the ertapenem bound to the cysteine. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

Ligplot automatically generates schematic diagrams of protein-ligand interactions mediated by hydrogen bonds and by hydrophobic contacts. This map of contacts will be used as a basis to look for a larger number of potential contacts between the two molecules in the collected ^{13}C-filtered-NOESY experiment. These very preliminary results are promising, as the bound-antibiotic seems to display several hydrophobic contacts with the protein, notably with Tyr^{403}, His^{421}, Trp^{425}, Gly^{438}, and Gly^{441}, in addition to hydrogen bonds with Ser^{439} and Ile^{443} in particular. These interactions could account for the difference in flexibility observed between the Ldt_{Bs} and Ldt_{fm} acylenzymes. Complementary experiments must also be led to investigate potential contacts between the aromatic ring of ertapenem and the protein.

Résumé de la partie III en français

Les LDts provenant de trois souches bactériennes différentes ont été étudiées durant cette thèse : la LDt de *Bacillus subtilis* (Ldt_{Bs}) tout d'abord, puis la LDt de *Enterococcus faecium* (Ldt_{fm}), et enfin l'une des LDts de *Mycobacterium tuberculosis* (Ldt_{Mt1}), dans une moindre mesure. Ces trois LDts ont été démontrées comme étant impliquées dans la formation de ponts 3→3 par transpeptidation dans leurs peptidoglycanes respectifs. L'objectif de ma thèse étant d'étudier le mécanisme de réaction d'acylation des LDts par les seuls β-lactames capables de les inhiber, les carbapénèmes, j'ai exploré par RMN trois des étapes de cette réaction : l'apoenzyme en absence d'antibiotique, la formation du complexe non-covalent, et enfin la formation de la liaison covalente entre la cystéine de l'enzyme et le groupement carbonyle du cycle β-lactame des carbapénèmes.

Pour Ldt_{Bs}, j'ai tout d'abord démontré la présence d'un réseau de liaisons hydrogène particulièrement stable sur une large gamme de pH (de pH 3.5 à 10.0), impliquant la triade catalytique Cys^{142}-His^{126}-Gly^{127}. La réorientation de la chaîne latérale de la cystéine active en absence de cation divalent, initialement proposée par les auteurs d'une structure cristallographique obtenue en présence de cadmium, a pu être démontrée expérimentalement.

La formation du complexe non covalent entre divers β-lactames et Ldt_{Bs}, rendue possible grâce à la mutation C142A, a démontré une interaction de très faible affinité ($K_D > 100$ mM et 37 mM pour l'ampicilline et l'imipénème, respectivement), induisant des changements structuraux négligeables, et une localisation très similaire de l'antibiotique au sein de la protéine pour les 2 familles de β-lactames. Ces résultats indiquent que la formation du complexe non-covalent n'est pas l'étape déterminante de la spécificité des carbapénèmes pour les LDts.

Enfin, la formation de l'acylenzyme, occasionnée par l'incubation en proportion 1 :1 de l'imipénème avec Ldt_{Bs}, induit d'importants réarrangements structuraux dans la protéine, ainsi qu'une significative augmentation de la dynamique. Trois régions de la protéine sont particulièrement affectées par la formation de la liaison covalente, et sont situées à proximité de la cystéine catalytique. La dynamique observée influe de manière drastique sur la qualité des spectres RMN, avec une disparition ou diminution de l'intensité des signaux, phénomène attribué à un échange conformationel se produisant à une échelle de temps similaire à celle des déplacements chimiques de la RMN. Des expériences de relaxation-dispersion ont été menées sur l'acylenzyme, ainsi que sur un autre adduit covalent, formé avec le 2-nitro-5-thiobenzoate (TNB), afin de comparer l'impact d'une liaison thioesther (Ldt_{Bs}-imipenem) et d'une liaison disulfure (Ldt_{Bs}-TNB) sur la dynamique. Cette seconde forme covalente nous a permis de modifier légèrement le régime d'échange, faisant réapparaitre quelques signaux sur les spectres RMN. L'échelle de temps de l'échange conformationel a été caractérisée dans les deux cas, avec une constante de vitesse d'environ d'environ 2 ms^{-1} pour les 3 régions affectées dans l'acylenzyme, et une constante de vitesse d'environ 6 ms^{-1} pour les trois régions semblables dans l'adduit contenant le TNB. Ces résultats indiquent que la formation d'une liaison covalente se trouve assurément à l'origine de la dynamique et du réarrangement structural, et que c'est cette étape qui est la clé dans la spécificité des carbapénèmes envers les LDts.

L'étude de la LDt de *Enterococcus faecium* comporte des similitudes avec celle de *Bacillus subtilis*. Tout d'abord, l'analyse thermodynamique et structurale de l'apoenzyme a révélé qu'un réseau de liaisons hydrogène similaire à celui observé dans le cas de Ldt_{Bs} ne pouvait pas être exclu. Ce réseau impliquerait les résidus : Cys^{442}-His^{421}-Asp^{422}. Cependant, la présence d'un résidu titrable (His^{440}) à proximité de Cys^{442} et His^{421} complique l'analyse de données de titration en fonction du pH pour ces deux résidus. La détermination des états de protonation dans la souche mutante H440A de Ldt_{fm} (en préparation) devrait nous permettre de lever les ambiguïtés d'interprétation qui persistent et d'affiner en conséquence la structure du site actif de Ldt_{fm} dans l'état apoenzyme. Ensuite, l'analyse du complexe non covalent de Ldt_{fm}-C442A avec des antibiotiques de trois familles différentes de β-lactames (péname, céphème et carbapénème) a montré que l'interaction est dans tous les cas de très faible affinité ($K_D \sim$ 50 mM pour la ceftriaxone et pour l'ertapénème), que le réarrangement structural est très faible à ce stade, et que les antibiotiques se lient non-covalemment dans des régions au niveau du site actif. L'étude sur ces deux LDts nous permet de conclure que la spécificité des LDts pour les carbapénèmes ne se fait pas au niveau de la formation du complexe de Michaelis-Menten. Enfin, la structure et la dynamique de l'acylenzyme Ldt_{fm}-ertapénème est en cours d'investigation. Les résultats obtenus à ce jour sont prometteurs quant à l'obtention de la première structure à haute résolution d'une LDt dans sa forme acylée. En effet, l'acylenzyme est beaucoup moins affectée par la dynamique que dans le cas de l'acylenzyme Ldt_{Bs}-imipénème, pour laquelle un large nombre de signaux RMN étaient indétectables. Des premiers contacts internucléaires ont déjà été observés entre l'ertapénème et le reste de la protéine Ldt_{fm} dans le cas du complexe covalent.

Enfin, la production de Ldt_{Mt1}, qui s'est révélée bien plus difficile qu'attendu, a récemment abouti à des conditions optimales pour une étude structurale par RMN. La suite du projet va donc inclure en priorité la résolution de la structure de Ldt_{Mt1}, puisqu'aucune des LDts de cette souche pathogène bactérienne n'est connue à ce jour. De plus, la comparaison du mécanisme d'acylation par les β-lactames à celui de l'acylation par les substrats naturels des LDts, les brins peptidiques du peptidoglycane, ouvrira de nouvelles voies pour la compréhension du mécanisme de transpeptidation et le design de nouvelles molécules inhibitrices thérapeutiques.

General conclusion and perspectives

In the present work we investigated by NMR the structure and dynamics of L,D-transpeptidases from three different bacterial strains, *B. subtilis* (Ldt_{Bs}), *E. faecium* (Ldt_{fm}), and to a lesser extent one of the five LDts from *M. tuberculosis* (Ldt_{Mt1}), as well as the thermodynamics of their non-covalent and covalent interaction with β-lactams, with a particular emphasis on carbapenems.

L,D-transpeptidases are inactivated by the formation of an acylenzyme in a two-step reaction involving the formation of a non-covalent complex prior to acylation of the active site cysteine of the enzyme by the carbonyl group of the carbapenem β-lactam ring. Three forms of the enzyme along this process were thoroughly investigated by NMR in the case of Ldt_{Bs} and Ldt_{fm}, namely the apoenzyme, the non-covalent complex, and the acylenzyme. X-ray structures were available for the two apoenzymes, yet they both contained artifactual ions within their active sites, thus disturbing the orientation of the residues involved in the catalytic process. In this context, the first step of the present PhD work was to investigate the structure of the ion-free apoenzymes by NMR, as the whole acylation reaction strongly depends on the initial conformation of the enzyme and to obtain a better resolution of the active site in a catalytically active form, as the acylation pathway may depend on the initial conformation of the enzyme.

Thermodynamical and structural NMR studies were first performed on Ldt_{Bs} apoenzyme and revealed the presence of a stable hydrogen bond network for the catalytic triad Cys^{142}-His^{126}-Gly^{127} over a surprising large pH range (from pH 3.5 to 10.0). Protonation states of ionizable residues involved in the active site were determined and led to the conclusion that Ldt_{Bs} exploits a neutral, hydrogen-bonded thiol-imidazole pair (SH/N) for catalysis, instead of the generally accepted thiolate-imidazolium (S^-/NH^+) ion pair found in the active site of some cysteine proteases. However, this mechanism needs to be confirmed in other LDts, in order to verify if this thiol-imidazole pair mechanism can apply to the entire LDt enzyme class or if it is a peculiarity of the *Bacillus subtilis* enzyme. The preliminary thermodynamical and structural studies performed on the Ldt_{fm} apoenzyme showed that the presence of a similar hydrogen bond network involving Cys^{442}-His^{421}-Asp^{422} may also occur, implying the same protonation states for each individual residue. Yet, the structure determination is still under the way, and the unambiguous determination of the Cys and His protonation states remains to be performed with the help of a comparison to the H440A mutant. In fact His^{440} has been identified as a likely origin for the hieratic behavior of Cys^{442} and His^{421} along titration experiments

Ldt_{Bs} and Ldt_{fm} non-covalent complexes were investigated with various β-lactam antibiotics from the penam, the carbapenem, and the cephem families. This study is inaccessible without the mutation of the catalytic cysteine into an alanine, thus preventing from the acylation reaction. The dissociation constants determined from the titration of Cys-into-Ala LDt mutants with β-lactams showed that the non-covalent interaction is very weak, with K_D values in the hundred of mM range for both enzymes. Furthermore, the analysis of chemical shift perturbations revealed that all antibiotics interact with the enzyme in the same region, located in the vicinity of the active site. These results demonstrate that the non-covalent interaction is not the key step towards β-lactam specificities. The latter rather arises from the last step of the reaction, namely the acylation step. This conclusion might also be relevant for the

LDt enzymes acylation by the natural nascent peptidoglycan substrates. One could picture that the peptidoglycan precursors might first non-covalently interact within the active site of LDts in a similar way to β-lactam antibiotics. The weak affinity and the small impact of the steric hindrance in the non-covalent interaction may compensate for the potential rigidity of the nascent peptidoglycan chains and thus facilitate the specificity of the chemical reaction afterwards in comparison to other enzymes.

The structure and the dynamics of the Ldt_{Bs}-imipenem and Ldt_{fm}-ertapenem acylenzymes were also investigated by NMR. In both cases, the acylation is associated with substantial conformational alteration and increased flexibility of the enzyme. In particular, a significant exchange contribution between several conformations was quantitatively characterized in the case of Ldt_{Bs} acylenzyme, with an exchange rate constant of 2 ms^{-1}. Exchange contributions were also detected in the Ldt_{fm} acylenzyme, while by far less extended, suggesting that the dynamics regime is variable depending on the enzyme. The increase in dynamics in the acylenzyme with respect to the apoenzyme does not seem to depend on the nature of the covalently attached molecule as the formation of a disulfide bond between a TNB molecule and Ldt_{Bs} induced conformational exchange with an exchange rate constant in the order of 6 ms^{-1}. These results suggest that the acylation of LDts with natural nascent peptidoglycan tetrapeptides may also influence the enzyme motional regime. Is the plasticity of the enzyme at the acylation stage a warranty for the L,D-transpeptidation efficiency ? Interestingly, an NMR study of one of the LMM PBPs involved in the D,D-transpeptidation pathway from *Streptomyces* (K15) was performed by our team and revealed a preexisting conformational exchange within the active site of the apoenzyme (unpublished results from Thomas Kern thesis). Taken together, these results suggest that the dynamic is likely to play a major role in both D,D and L,D-transpeptidation pathways.

Finally, the biochemical results recently obtained for Ldt_{Mt1} are very promising for further NMR studies on this enzyme. The structure and dynamics of Ldt_{Mt1} enzyme will be explored by NMR to assess the role played by conformational flexibility in the efficiency of the acylation reaction and to further decipher the inhibition mechanism of LDts by carbapenems and related β-lactam classes. As this enzyme is a promising target for the treatment of multi-resistant *M. tuberculosis*, the obtention of the first LDt structure from this pathogenic strain might open the way for the design of new antibiotics. Although drug development is not an objective of the current project that was recently funded by the ANR, the knowledge base built on Ldt_{Mt1} will clearly be an asset for rational design of L,D-transpeptidase inhibitors. In parallel to NMR and kinetics studies, one of the partner group from the ANR project (Mélanie Ethève-Quelquejeu team) is working on the synthesis of various carbapenems in order to decipher the role of the side-chains of the drugs for binding. The possibilities that the side-chain promotes or prevents drug binding, and the proper orientation of the β-lactam ring will be considered at the stage of the nucleophilic attack by the catalytic cysteine. In addition, the comparison of the inhibitor and the physiological substrates will be studied and will enable to determine how carbapenems mimic the peptidoglycan precursors and to understand the basis for the stability of the drug-Ldt_{Mt1} adduct, which is essential for antibacterial activity.

Another perspective of the project will be to investigate the chemical step of the reaction by analyzing both the reactivity of the β-lactam rings of various antibiotic families and the reactivity of the active cysteine in the environment of the catalytic cavity. Quantum chemistry will be used to generate

working hypotheses about the path of the acylation reaction and the contribution of the carbapenem ring to substrate specificity. These hypotheses will be tested by structure-assisted site-directed mutagenesis of Ldt_{Mt1} and by testing different β-lactams. However, the quantum chemistry needs to deal with an accurate acylenzyme structure. As the crystallization of acylenzymes failed to-date (personal communication from Michel Arthur), it is thus required to solve the NMR structure of one of the LDt-acylenzymes with a good resolution at the active site and a proper orientation of the carbepenem. The 3D NMR structure of Ldt_{fm}-ertapenem acylenzyme that is under way represents a significant progress towards this goal. The diminished impact of the acylenzyme flexibility in comparison to Ldt_{Bs} and the already detected intermolecular contacts between the enzyme and the ertapenem antibiotic sound really promising in this perspective. Ldt_{Mt1} may also turn out as a particularly interesting target as clear differences in the kinetics of the acylation have been shown within the carbapenem and cephalosporin β-lactam families.

The combination of complementary techniques and expertise available in the different research groups implied in the present ANR-funded project provide an ideal environment to answer the challenging and numerous questions about the acylation reaction of LDts, with respect to either their inactivation by β-lactams of the carbapenem class, or the peptidoglycan transpeptidation mechanism. The present work has provided new insights into the acylation mechanism of LDts with carbapenems through the NMR studies of Ldt_{Bs} and Ldt_{fm}, and the further investigations on Ldt_{fm} and Ldt_{Mt1} will hopefully bring the anticipated results, leading to a generalized mechanism for the entire enzyme family.

Bibliographie

Abraham, E. P. and Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. *Rev Infect Dis*, 10(4) :677–678.

Abraham, E. P. and Newton, G. G. (1961). New penicillins, cephalosporin C, and penicillinase. *Endeavour*, 20 :92–100.

Altieri, A. S. and Byrd, R. A. (2004). Automation of NMR structure determination of proteins. *Curr Opin Struct Biol*, 14(5) :547–553.

Ambler, R. P. (1980). The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci*, 289(1036) :321–331.

Anderson, J. S., Matsuhashi, M., Haskin, M. A., and Strominger, J. L. (1965). Lipid-phosphoacetylmuramyl-pentapeptide and lipid-phosphodisaccharide-pentapeptide : presumed membrane transport intermediates in cell wall synthesis. *Proc Natl Acad Sci U S A*, 53 :881–889.

André, I., Linse, S., and Mulder, F. A. A. (2007). Residue-specific pKa determination of lysine and arginine side chains by indirect ^{15}N and ^{13}C NMR spectroscopy : application to apo calmodulin. *J Am Chem Soc*, 129(51) :15805–15813.

Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M. J., Gans, P., and Boisbouvier, J. (2012). An optimized isotopic labelling strategy of isoleucine-γ2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem Commun (Camb)*, 48(10) :1434–1436.

Bachovchin, W. W. (1986). ^{15}N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases : evidence for a moving histidine mechanism. *Biochemistry*, 25(23) :7751–7759.

Baddiley, J. (1972). Teichoic acids in cell walls and membranes of bacteria. *Essays Biochem*, 8 :35–77.

Balaji, V., Jeremiah, S. S., and Baliga, P. R. (2011). Polymyxins : Antimicrobial susceptibility concerns and therapeutic options. *Indian J Med Microbiol*, 29(3) :230–242.

Bardiaux, B., Malliavin, T., and Nilges, M. (2012). ARIA for solution and solid-state NMR. *Methods Mol Biol*, 831 :453–483.

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., and Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev*, 32(2) :168–207.

Barrett, D., Wang, T.-S. A., Yuan, Y., Zhang, Y., Kahne, D., and Walker, S. (2007). Analysis of glycan polymers produced by peptidoglycan glycosyltransferases. *J Biol Chem*, 282(44) :31964–31971.

Bartlett, G. J., Porter, C. T., Borkakoti, N., and Thornton, J. M. (2002). Analysis of catalytic residues in enzyme active sites. *J Mol Biol*, 324(1) :105–121.

Bateman, A. and Bycroft, M. (2000). The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol*, 299 :1113–1119.

Bax, A., Clore, G. M., and Gron (1990). ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.*, 88 :425–431.

Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Crystal structure of the eukaryotic ribosome. *Science*, 330 :1203–1209.

Benson, T. E., Walsh, C. T., and Hogle, J. M. (1996). The structure of the substrate-free form of MurB, an essential enzyme for the synthesis of bacterial cell walls. *Structure*, 4(1) :47–54.

Bentley, R. (2009). Different roads to discovery ; prontosil (hence sulfa drugs) and penicillin (hence beta-lactams). *J Ind Microbiol Biotechnol*, 36(6) :775–786.

Bertsche, U., Breukink, E., Kast, T., and Vollmer, W. (2005). In vitro murein peptidoglycan synthesis by dimers of the bifunctional transglycosylase-transpeptidase PBP1b from *Escherichia coli*. *J Biol Chem*, 280(45) :38096–38101.

Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A., and Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol*, 43(3) :717–731.

Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007). Evaluating protein structures determined by structural genomics consortia. *Proteins*, 66(4) :778–795.

Bhattacharyya, B., Kapoor, S., and Panda, D. (2010). Fluorescence spectroscopic methods to analyze drug-tubulin interactions. *Methods Cell Biol*, 95 :301–329.

Biarrotte-Sorin, S., Hugonnet, J.-E., Delfosse, V., Mainardi, J.-L., Gutmann, L., Arthur, M., and Mayer, C. (2006). Crystal structure of a novel β-lactam-insensitive peptidoglycan transpeptidase. *J Mol Biol*, 359(3) :533–538.

Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z. S. (2006). B. subtilis ykud protein at 2.0 a resolution : insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. *Proteins*, 62(1) :144–151.

Blasco, B., Pisabarro, A. G., and de Pedro, M. A. (1988). Peptidoglycan biosynthesis in stationary-phase cells of escherichia coli. *J Bacteriol*, 170(11) :5224–5228.

Blomberg, F., Maurer, W., and Rüterjans, H. (1977). Nuclear magnetic resonance investigation of ^{15}N-labeled histidine in aqueous solution. *J Am Chem Soc*, 99(25) :8149–8159.

Boneca, I. G., Huang, Z. H., Gage, D. A., and Tomasz, A. (2000). Characterization of *Staphylococcus aureus* cell wall glycan strands, evidence for a new beta-N-acetylglucosaminidase activity. *J Biol Chem*, 275(14) :9910–9918.

Bos, M. P., Robert, V., and Tommassen, J. (2007). Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol*, 61 :191–214.

Bougault, C., Heidiger, S., and Simorre, J.-P. (2012). *Bacterial Glycomics. Current Research, Technology and Applications. Chapter 9 : "Solid-state NMR of the Bacterial Cell Wall"*. Reid, C. W. and Twine, S. M. and Reid, A. N.

Bouhss, A., Crouvoisier, M., Blanot, D., and Mengin-Lecreulx, D. (2004). Purification and characterization of the bacterial mray translocase catalyzing the first membrane step of peptidoglycan biosynthesis. *J Biol Chem*, 279(29) :29974–29980.

Bouhss, A., Trunkfield, A. E., Bugg, T. D. H., and Mengin-Lecreulx, D. (2008). The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev*, 32(2) :208–233.

Breukels, V., Konijnenberg, A., Nabuurs, S. M., Doreleijers, J. F., Kovalevskaya, N. V., and Vuister, G. W. (2011). Overview on the use of NMR to examine protein structure. *Curr Protoc Protein Sci*, Chapter 17 :Unit17.5.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Crystallography & NMR system : A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, 54(Pt 5) :905–921.

Bush, K. (2004). Antibacterial drug discovery in the 21st century. *Clin Microbiol Infect*, 10 Suppl 4 :10–17.

Carlomagno, T. (2005). Ligand-target interactions : what can we learn from NMR ? *Annu Rev Biophys Biomol Struct*, 34 :245–266.

Carr, H. Y. and Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, 94 :630–638.

Cavanagh, J., Frairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1995). *Protein NMR Spectroscopy. Principles and Practice.* Academic press.

Changeux, J.-P. and Edelstein, S. (2011). Conformational selection or induced fit ? 50 years of debate resolved. *F1000 Biol Rep*, 3 :19.

Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O., and Bobik, T. A. (2008). Bacterial microcompartments : their properties and paradoxes. *Bioessays*, 30(11-12) :1084–1095.

Chiu, W. (1993). What does electron cryomicroscopy provide that X-ray crystallography and NMR spectroscopy cannot ? *Annu Rev Biophys Biomol Struct*, 22 :233–255.

Chou, J. J., Gaemers, S., Howder, B., Louis, J. M., and Bax, A. (2001). A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR*, 21(4) :377–382.

Clore, G. M., Starich, M. R., and Gronenborn, A. M. (1998). Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc*, 120(40) :10571–10572.

Cole, R. and Loria, J. P. (2003). Fast-modelfree : a program for rapid automated analysis of solution nmr spin-relaxation data. *J Biomol NMR*, 26(3) :203–213.

Constantine, K. L., Mueller, L., Goldfarb, V., Wittekind, M., Metzler, W. J., Yanchunas, Jr, J., Robertson, J. G., Malley, M. F., Friedrichs, M. S., and Farmer, 2nd, B. (1997). Characterization of NADP+ binding to perdeuterated MurB : backbone atom NMR assignments and chemical-shift changes. *J Mol Biol*, 267(5) :1223–1246.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, 13(3) :289–302.

Craik, D. J. and Wilce, J. A. (1997). Studies of protein-ligand interactions by NMR. *Methods Mol Biol*, 60 :195–232.

Croke, R. L., Patil, S. M., Quevreaux, J., Kendall, D. A., and Alexandrescu, A. T. (2011). NMR determination of pka values in α-synuclein. *Protein Sci*, 20(2) :256–269.

Daffé, M. and Reyrat, J.-M. (2008). *The mycobacterial cell envelope*. Washington : ASM Press.

Dai, H. J., Parker, C. N., and Bao, J. J. (2002). Characterization and inhibition study of MurA enzyme by capillary electrophoresis. *J Chromatogr B Analyt Technol Biomed Life Sci*, 766(1) :123–132.

Day, R. M., Thalhauser, C. J., Sudmeier, J. L., Vincent, M. P., Torchilin, E. V., Sanford, D. G., Bachovchin, C. W., and Bachovchin, W. W. (2003). Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN' by NMR. *Protein Sci*, 12(4) :794–810.

De Lencastre, H. and Tomasz, A. (2002). From ecological reservoir to disease : the nasopharynx, day-care centres and drug-resistant clones of *Streptococcus pneumoniae*. *J Antimicrob Chemother*, 50 Suppl S2 :75–81.

de Vries, S. J., van Dijk, M., and Bonvin, A. M. J. J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc*, 5(5) :883–897.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). Nmrpipe : a multidimensional spectral processing system based on unix pipes. *J Biomol NMR*, 6(3) :277–293.

Delaglio, F., Kontaxis, G., and Bax, A. (2000). Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J Am Chem Soc*, 122 :2142–2143.

Demain, A. L. and Elander, R. P. (1999). The beta-lactam antibiotics : past, present, and future. *Antonie Van Leeuwenhoek*, 75(1-2) :5–19.

Diehl, C., EngstrÃ¶m, O., Delaine, T., HÃ¥kansson, M., Genheden, S., Modig, K., Leffler, H., Ryde, U., Nilsson, U. J., and Akke, M. (2010). Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc*, 132(41) :14577–14589.

Dmitriev, B. A., Ehlers, S., and Rietschel, E. T. (1999). Layered murein revisited : a fundamentally new concept of bacterial cell wall structure, biogenesis and function. *Med Microbiol Immunol*, 187(3) :173–181.

Dmitriev, B. A., Ehlers, S., Rietschel, E. T., and Brennan, P. J. (2000). Molecular mechanics of the mycobacterial cell wall : from horizontal layers to vertical scaffolds. *Int J Med Microbiol*, 290(3) :251–258.

Dmitriev, B. A., Toukach, F. V., Schaper, K.-J., Holst, O., Rietschel, E. T., and Ehlers, S. (2003). Tertiary structure of bacterial murein : the scaffold model. *J Bacteriol*, 185(11) :3458–3468.

Domagk, G. (1957). Twenty-five years of sulfonamide therapy. *Ann N Y Acad Sci*, 69(3) :380–384.

Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). Haddock : a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, 125(7) :1731–1737.

Doreleijers, J. F., Vranken, W. F., Schulte, C., Markley, J. L., Ulrich, E. L., Vriend, G., and Vuister, G. W. (2012). NRG-CING : integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic Acids Res*, 40(Database issue) :D519–D524.

Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR*, 16(1) :23–28.

Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001). A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J Biomol NMR*, 20(3) :223–231.

Drawz, S. M. and Bonomo, R. A. (2010). Three decades of beta-lactamase inhibitors. *Clin Microbiol Rev*, 23(1) :160–201.

Drlica, K., Hiasa, H., Kerns, R., Malik, M., Mustaev, A., and Zhao, X. (2009). Quinolones : action and resistance updated. *Curr Top Med Chem*, 9(11) :981–998.

Dubée, V., Arthur, M., Fief, H., Triboulet, S., Mainardi, J.-L., Gutmann, L., Sollogoub, M., Rice, L. B., Ethève-Quelquejeu, M., and Hugonnet, J.-E. (2012a). Kinetic analysis of *Enterococcus faecium* L,D-transpeptidase inactivation by carbapenems. *Antimicrob Agents Chemother*, 56(6) :3409–3412.

Dubée, V., Triboulet, S., Mainardi, J.-L., Ethève-Quelquejeu, M., Gutmann, L., Marie, A., Dubost, L., Hugonnet, J.-E., and Arthur, M. (2012b). Inactivation of *Mycobacterium tuberculosis* L,D-transpeptidase Ldt_{Mt1} by carbapenems and cephalosporins. *Antimicrob Agents Chemother*, 56(8) :4189–4195.

Duff, Jr, M. R., Grubbs, J., and Howell, E. E. (2011). Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J Vis Exp*, 55(55) :1–4.

Dutta, R., Qin, L., and Inouye, M. (1999). Histidine kinases : diversity of domain organization. *Mol Microbiol*, 34(4) :633–640.

Everill, P., Sudmeier, J. L., and Bachovchin, W. W. (2012). Direct NMR observation and pka determination of the asp102 side chain in a serine protease. *J Am Chem Soc*, 134(4) :2348–2354.

Fairbrother, R. W. and Taylor, G. (1961). Sodium methicillin in routine therapy. *Lancet*, 1(7175) :473–476.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, 33(19) :5984–6003.

Feuerstein, S., Plevin, M. J., Willbold, D., and Brutscher, B. (2012). iHADAMAC : a complementary tool for sequential resonance assignment of globular and highly disordered proteins. *J Magn Reson*, 214(1) :329–334.

Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002). NMR analysis of a 900K GroEL GroES complex. *Nature*, 418(6894) :207–211.

Fielding, L. (2007). NMR methods for the determination of protein-ligand dissociation constants. *Prog. NMR Spectrosc.*, 51(1) :219–242.

Fischer, M. W. F., Majumdar, A., and Zuiderweg, E. R. P. (1998). *Protein NMR relaxation : theory, applications and outlook.* Progress in nuclear magnetic resonance spectroscopy 33(4) : 207-272.

Floss, H. G. and Yu, T.-W. (2005). Rifamycin-mode of action, resistance, and biosynthesis. *Chem Rev*, 105(2) :621–632.

Fontana, R., Aldegheri, M., Ligozzi, M., Lopez, H., Sucari, A., and Satta, G. (1994). Overproduction of a low-affinity penicillin-binding protein and high-level ampicillin resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother*, 38(9) :1980–1983.

Fontana, R., Cerini, R., Longoni, P., Grossato, A., and Canepari, P. (1983). Identification of a streptococcal penicillin-binding protein that reacts very slowly with penicillin. *J Bacteriol*, 155(3) :1343–1350.

Frère, J. M. (1995). Beta-lactamases and bacterial resistance to antibiotics. *Mol Microbiol*, 16(3) :385–395.

Frère, J. M., Duez, C., Ghuysen, J. M., and Vandekerkhove, J. (1976). Occurrence of a serine residue in the penicillin-binding site of the exocellular dd-carboxy-peptidase-transpeptidase from streptomyces r61. *FEBS Lett*, 70(1) :257–260.

Frère, J. M., Ghuysen, J. M., and Perkins, H. R. (1975). Interaction between the exocellular dd-carboxypeptidase-transpeptidase from streptomyces r61, substrate and beta-lactam antibiotics. a choice of models. *Eur J Biochem*, 57(2) :353–359.

Freyssingeas, E., Nallet, F., and Roux, D. (1996). measurement of the membrane flexibility in lamellar and "sponge" phases of the $c_{12}e_5$/hexanol/water system. *Langmuir*, 12 :6028–6035.

Gabel, F., Bicout, D., Lehnert, U., Tehei, M., Weik, M., and Zaccai, G. (2002). Protein dynamics studied by neutron scattering. *Quaterly Reviews of Biophysics*, 35(4) :327–367.

Gan, L., Chen, S., and Jensen, G. J. (2008). Molecular organization of gram-negative peptidoglycan. *Proc Natl Acad Sci U S A*, 105(48) :18953–18957.

Gans, P., Hamelin, O., Sounier, R., Ayala, I., DurÃ¡, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J., and Boisbouvier, J. (2010). Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl*, 49(11) :1958–1962.

García de la Torre, J., Huertas, M. L., and Carrasco, B. (2000). HYDRONMR : prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*, 147(1) :138–146.

Ghuysen, J. M., Leyh-Bouille, M., Campbell, J. N., Moreno, R., Frére, J. M., Duez, C., Nieto, M., and Perkins, H. R. (1973). Structure of the wall peptidoglycan of streptomyces r39 and the specificity profile of its exocellular dd-carboxypeptidase–transpeptidase for peptide acceptors. *Biochemistry*, 12(7) :1243–1251.

Glauner, B. (1988). Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal Biochem*, 172(2) :451–464.

Glauner, B. and Höltje, J. V. (1990). Growth pattern of the murein sacculus of *Escherichia coli*. *J Biol Chem*, 265(31) :18988–18996.

Glauner, B., Höltje, J. V., and Schwarz, U. (1988). The composition of the murein of *Escherichia coli*. *J Biol Chem*, 263(21) :10088–10095.

Goffin, C. and Ghuysen, J. M. (1998). Multimodular penicillin-binding proteins : an enigmatic family of orthologs and paralogs. *Microbiol Mol Biol Rev*, 62(4) :1079–1093.

Gordon, E., Mouz, N., Duée, E., and Dideberg, O. (2000). The crystal structure of the penicillin-binding protein 2x from streptococcus pneumoniae and its acyl-enzyme form : implication in drug resistance. *J Mol Biol*, 299(2) :477–485.

Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2009). A summary of the measured pk values of the ionizable groups in folded proteins. *Protein Sci*, 18(1) :247–251.

Grishaev, A., Tugarinov, V., Kay, L. E., Trewhella, J., and Bax, A. (2008). Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR*, 40(2) :95–106.

Grzesiek, S. and Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J Biomol NMR*, 3(2) :185–204.

Guerry, P. and Herrmann, T. (2012). Comprehensive automation for NMR structure determination of proteins. *Methods Mol Biol*, 831 :429–451.

Güntert, P. (1997). Calculating protein structures from NMR data. *Methods Mol Biol*, 60 :157–194.

Güntert, P. (2003). Automated NMR protein structure calculation. *Prog. NMR Spectrosc.*, 43 :105–125.

Gupta, R., Lavollay, M., Mainardi, J.-L., Arthur, M., Bishai, W. R., and Lamichhane, G. (2010). The *Mycobacterium tuberculosis* protein Ldt_{Mt2} is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. *Nat Med*, 16(4) :466–469.

Ha, S., Walker, D., Shi, Y., and Walker, S. (2000). The 1.9 a crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci*, 9(6) :1045–1052.

Hakenbeck, R. and Coyette, J. (1998). Resistant penicillin-binding proteins. *Cell Mol Life Sci*, 54(4) :332–340.

Hammes, G. G., Chang, Y.-C., and Oas, T. G. (2009). Conformational selection or induced fit : a flux description of reaction mechanism. *Proc Natl Acad Sci U S A*, 106(33) :13737–13741.

Hansen, M. R., Mueller, L., and Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol*, 5(12) :1065–1074.

Hass, M. A. S., Hansen, D. F., Christensen, H. E. M., Led, J. J., and Kay, L. E. (2008). Characterization of conformational exchange of a histidine side chain : protonation, rotamerization, and tautomerization of his61 in plastocyanin from anabaena variabilis. *J Am Chem Soc*, 130(26) :8460–8470.

Hayhurst, E. J., Kailas, L., Hobbs, J. K., and Foster, S. J. (2008). Cell wall peptidoglycan architecture in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, 105(38) :14603–14608.

Healy, V. L., Lessard, I. A., Roper, D. I., Knox, J. R., and Walsh, C. T. (2000). Vancomycin resistance in enterococci : reprogramming of the D-ala-D-ala ligases in bacterial peptidoglycan biosynthesis. *Chem Biol*, 7(5) :R109–R119.

Helfand, M. S. and Bonomo, R. A. (2003). Beta-lactamases : a survey of protein diversity. *Curr Drug Targets Infect Disord*, 3(1) :9–23.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated noe-identification in the noesy spectra using the new software atnos. *J Biomol NMR*, 24(3) :171–189.

Herzberg, O. and Moult, J. (1987). Bacterial resistance to beta-lactam antibiotics : crystal structure of beta-lactamase from *Staphylococcus aureus* pc1 at 2.5 Å resolution. *Science*, 236(4802) :694–701.

Höltje, J. V. (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiol Mol Biol Rev*, 62(1) :181–203.

Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. (2003). Crystal structure of the MurG :UDP-GlcNac complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci U S A*, 100(3) :845–849.

Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S., and Palzkill, T. (1996). Amino acid sequence determinants of beta-lactamase structure and activity. *J Mol Biol*, 258(4) :688–703.

Hugonnet, J.-E., Tremblay, L. W., Boshoff, H. I., Barry, 3rd, C. E., and Blanchard, J. S. (2009). Meropenem-clavulanate is effective against extensively drug-resistant *Mycobacterium tuberculosis*. *Science*, 323(5918) :1215–1218.

Hus, J. C., Marion, D., and Blackledge, M. (2001). Determination of protein backbone structure using only residual dipolar couplings. *J Am Chem Soc*, 123(7) :1541–1542.

Ikeda, M., Sato, T., Wachi, M., Jung, H. K., Ishino, F., Kobayashi, Y., and Matsuhashi, M. (1989). Structural similarity among *Escherichia coli* FtsW and RodA proteins and *Bacillus subtilis* SpoVE protein, which function in cell division, cell elongation, and spore formation, respectively. *J Bacteriol*, 171(11) :6375–6378.

Izoré, T., Job, V., and Dessen, A. (2011). Biogenesis, regulation, and targeting of the type III secretion system. *Structure*, 19(5) :603–612.

Jabes, D. (2011). The antibiotic r&d pipeline : an update. *Curr Opin Microbiol*, 14(5) :564–569.

Jain, N. U. (2009). Use of residual dipolar couplings in structural analysis of protein-ligand complexes by solution NMR spectroscopy. *Methods Mol Biol*, 544 :231–252.

Jana, S. and Deb, J. K. (2006). Molecular understanding of aminoglycoside action and resistance. *Appl Microbiol Biotechnol*, 70(2) :140–150.

Jelesarov, I. and Bosshard, H. R. (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit*, 12(1) :3–18.

Jensen, M. R., Hass, M. A. S., Hansen, D. F., and Led, J. J. (2007). Investigating metal-binding in proteins by nuclear magnetic resonance. *Cell Mol Life Sci*, 64(9) :1085–1104.

Johnson, E. S. (2004). Protein modification by SUMO. *Annu Rev Biochem*, 73 :355–382.

Kahan, J. S., Kahan, F. M., Goegelman, R., Currie, S. A., Jackson, M., Stapley, E. O., Miller, T. W., Miller, A. K., Hendlin, D., Mochales, S., Hernandez, S., Woodruff, H. B., and Birnbaum, J. (1979). Thienamycin, a new beta-lactam antibiotic. i. discovery, taxonomy, isolation and physical properties. *J Antibiot (Tokyo)*, 32(1) :1–12.

Kanelis, V., Forman-Kay, J. D., and Kay, L. E. (2001). Multidimensional NMR methods for protein structure determination. *IUBMB Life*, 52(6) :291–302.

Kannan, K. and Mankin, A. S. (2011). Macrolide antibiotics in the ribosome exit tunnel : species-specific binding and action. *Ann N Y Acad Sci*, 1241 :33–47.

Karplus, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.*, 30 :11–15.

Karplus, M. (1963). Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.*, 18 :2870–2871.

Kay, L. E. (1998). Protein dynamics from NMR. *Nat Struct Biol*, 5 Suppl :513–517.

Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy : application to staphylococcal nuclease. *Biochemistry*, 28(23) :8972–8979.

Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. *J Magn Reson*, 101(Series B) :333–337.

Kern, T., Hediger, S., Müller, P., Giustini, C., Joris, B., Bougault, C., Vollmer, W., and Simorre, J.-P. (2008). Toward the characterization of peptidoglycan structure and protein-peptidoglycan interactions by solid-state NMR spectroscopy. *J Am Chem Soc*, 130(17) :5618–5619.

Khosla, R., Verma, D. D., Kapur, A., Aruna, R. V., and Khanna, N. (1999). Streptogramins : a new class of antibiotics. *Indian J Med Sci*, 53(3) :111–119.

Kim, J., Mao, J., and Gunner, M. R. (2005). Are acidic and basic groups in buried proteins predicted to be ionized ? *J Mol Biol*, 348(5) :1283–1298.

Klare, I., Rodloff, A. C., Wagner, J., Witte, W., and Hakenbeck, R. (1992). Overproduction of a penicillin-binding protein is not the only mechanism of penicillin resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother*, 36(4) :783–787.

Kleckner, I. R. and Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta*, 1814(8) :942–968.

Kneller, J. M., Lu, M., and Bracken, C. (2002). An effective method for the discrimination of motional anisotropy and chemical exchange. *J Am Chem Soc*, 124(9) :1852–1853.

Kodama, T., Takamatsu, H., Asai, K., Ogasawara, N., Sadaie, Y., and Watabe, K. (2000). Synthesis and characterization of the spore proteins of *Bacillus subtilis* YdhD, YkuD, and YkvP, which carry a motif conserved among cell wall binding proteins. *J Biochem*, 128(4) :655–663.

Kohlrausch, U., Wientjes, F. B., and Höltje, J. V. (1989). Determination of murein precursors during the cell cycle of *Escherichia coli*. *J Gen Microbiol*, 135(6) :1499–1506.

Komeili, A. (2012). Molecular mechanisms of compartmentalization and biomineralization in magnetotactic bacteria. *FEMS Microbiol Rev*, 36(1) :232–255.

Kong, K.-F., Schneper, L., and Mathee, K. (2010). Beta-lactam antibiotics : from antibiosis to resistance and bacteriology. *APMIS*, 118(1) :1–36.

Kumar, P., Arora, K., Lloyd, J. R., Lee, I. Y., Nair, V., Fischer, E., Boshoff, H. I. M., and Barry, 3rd, C. E. (2012). Meropenem inhibits D,D-carboxypeptidase activity in *Mycobacterium tuberculosis*. *Mol Microbiol*, 86(2) :367–381.

Kupce, E. and Freeman, R. (2003). Frequency-domain Hadamard spectroscopy. *J Magn Reson*, 162(1) :158–165.

Kuszewski, J., Schwieters, C. D., Garrett, D. S., Byrd, R. A., Tjandra, N., and Clore, G. M. (2004). Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments. *J Am Chem Soc*, 126(20) :6258–6273.

Labischinski, H., Barnickel, G., Bradaczek, H., and Giesbrecht, P. (1979). On the secondary and tertiary structure of murein. low and medium-angle x-ray evidence against chitin-based conformations of bacterial peptidoglycan. *Eur J Biochem*, 95(1) :147–155.

Labischinski, H., Goodell, E. W., Goodell, A., and Hochberg, M. L. (1991). Direct proof of a ”more-than-single-layered” peptidoglycan architecture of *Escherichia coli* W7 : a neutron small-angle scattering study. *J Bacteriol*, 173(2) :751–756.

Laraki, N., Galleni, M., Thamm, I., Riccio, M. L., Amicosante, G., FrÃ¨re, J. M., and Rossolini, G. M. (1999). Structure of in31, a blaimp-containing pseudomonas aeruginosa integron phyletically related to in5, which carries an unusual array of gene cassettes. *Antimicrob Agents Chemother*, 43(4) :890–901.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). Aqua and procheck-NMR : programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, 8(4) :477–486.

Lauretti, L., Riccio, M. L., Mazzariol, A., Cornaglia, G., Amicosante, G., Fontana, R., and Rossolini, G. M. (1999). Cloning and characterization of blavim, a new integron-borne metallo-beta-lactamase gene from a pseudomonas aeruginosa clinical isolate. *Antimicrob Agents Chemother*, 43(7) :1584–1590.

Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Riegel, P., Gutmann, L., and Mainardi, J.-L. (2009). The beta-lactam-sensitive D,D-carboxypeptidase activity of PBP4 controls the L,D and D,D transpeptidation pathways in *Corynebacterium jeikeium*. *Mol Microbiol*, 74(3) :650–661.

Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., Blanot, D., Gutmann, L., and Mainardi, J.-L. (2008). The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L,D-transpeptidation. *J Bacteriol*, 190(12) :4360–4366.

Lavollay, M., Fourgeaud, M., Herrmann, J.-L., Dubost, L., Marie, A., Gutmann, L., Arthur, M., and Mainardi, J.-L. (2011). The peptidoglycan of *Mycobacterium abscessus* is predominantly cross-linked by L,D-transpeptidases. *J Bacteriol*, 193(3) :778–782.

Leclercq, R. and Courvalin, P. (1991). Bacterial resistance to macrolide, lincosamide, and streptogramin antibiotics by target modification. *Antimicrob Agents Chemother*, 35(7) :1267–1272.

Lecoq, L., Bougault, C., Kern, T., Hugonnet, J.-E., Veckerlé, C., Pessey, O., Arthur, M., and Simorre, J.-P. (2012). Backbone and side-chain ^{1}H, ^{15}N and ^{13}C assignment of apo- and imipenem-acylated L,D-transpeptidase from *Bacillus subtilis*. *Biomol NMR Assign*.

Lescop, E., Rasia, R., and Brutscher, B. (2008). Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J Am Chem Soc*, 130(15) :5014–5015.

Lescop, E., Schanda, P., and Brutscher, B. (2007). A set of best triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson*, 187(1) :163–169.

Leszczynski, J. F. and Rose, G. D. (1986). Loops in globular proteins : a novel category of secondary structure. *Science*, 234(4778) :849–855.

Lewis, S. D., Johnson, F. A., and Shafer, J. A. (1981). Effect of cysteine-25 on the ionization of histidine-159 in papain as determined by proton nuclear magnetic resonance spectroscopy. evidence for a his-159–cys-25 ion pair and its possible role in catalysis. *Biochemistry*, 20(1) :48–51.

Li, S. and Hong, M. (2011). Protonation, tautomerization, and rotameric structure of histidine : a comprehensive study by magic-angle-spinning solid-state NMR. *J Am Chem Soc*, 133(5) :1534–1544.

Lim, D. and Strynadka, N. C. J. (2002). Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Nat Struct Biol*, 9(11) :870–876.

Lipari, G. and Szabo, A. (1982a). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*, 104(17) :4546–4559.

Lipari, G. and Szabo, A. (1982b). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc*, 104(17) :4559–4570.

Lipsitz, R. S. and Tjandra, N. (2004). Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct*, 33 :387–413.

Livermore, D. M. (2012). Fourteen years in resistance. *Int J Antimicrob Agents*, 39(4) :283–294.

Long, K. S., Hansen, L. H., Jakobsen, L., and Vester, B. (2006). Interaction of pleuromutilin derivatives with the ribosomal peptidyl transferase center. *Antimicrob Agents Chemother*, 50(4) :1458–1462.

Loria, J. P., Rance, M., and Palmer, 3rd, A. (1999a). A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc*, 121 :2231–2232.

Loria, J. P., Rance, M., and Palmer, 3rd, A. (1999b). A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J Biomol NMR*, 15(2) :151–155.

Lovering, A. L., de Castro, L. H., Lim, D., and Strynadka, N. C. J. (2007). Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science*, 315(5817) :1402–1405.

Lovering, A. L., Gretes, M., and Strynadka, N. C. J. (2008). Structural details of the glycosyltransferase step of peptidoglycan assembly. *Curr Opin Struct Biol*, 18(5) :534–543.

Lovering, A. L., Safadi, S. S., and Strynadka, N. C. J. (2012). Structural perspective of peptidoglycan biosynthesis and assembly. *Annu Rev Biochem*, 81 :451–478.

Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O., and Dessen, A. (2006). Penicillin binding proteins : key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiol Rev*, 30(5) :673–691.

Macheboeuf, P., Piuzzi, M., Finet, S., Bontems, F., Pérez, J., Dessen, A., and Vachette, P. (2011). Solution x-ray scattering study of a full-length class a penicillin-binding protein. *Biochem Biophys Res Commun*, 405(1) :107–111.

Magnet, S., Arbeloa, A., Mainardi, J.-L., Hugonnet, J.-E., Fourgeaud, M., Dubost, L., Marie, A., Delfosse, V., Mayer, C., Rice, L. B., and Arthur, M. (2007a). Specificity of L,D-transpeptidases from Gram-positive bacteria producing different peptidoglycan chemotypes. *J Biol Chem*, 282(18) :13151–13159.

Magnet, S., Bellais, S., Dubost, L., Fourgeaud, M., Mainardi, J.-L., Petit-Frère, S., Marie, A., Mengin-Lecreulx, D., Arthur, M., and Gutmann, L. (2007b). Identification of the L,D-transpeptidases responsible for attachment of the braun lipoprotein to *Escherichia coli* peptidoglycan. *J Bacteriol*, 189(10) :3927–3931.

Magnet, S., Dubost, L., Marie, A., Arthur, M., and Gutmann, L. (2008). Identification of the L,D-transpeptidases for peptidoglycan cross-linking in *Escherichia coli*. *J Bacteriol*, 190(13) :4782–4785.

Mahapatra, S., Crick, D. C., and Brennan, P. J. (2000). Comparison of the UDP-N-acetylmuramate :L-alanine ligase enzymes from *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *J Bacteriol*, 182(23) :6827–6830.

Mainardi, J.-L., Fourgeaud, M., Hugonnet, J.-E., Dubost, L., Brouard, J.-P., Ouazzani, J., Rice, L. B., Gutmann, L., and Arthur, M. (2005). A novel peptidoglycan cross-linking enzyme for a β-lactam-resistant transpeptidation pathway. *J Biol Chem*, 280(46) :38146–38152.

Mainardi, J. L., Legrand, R., Arthur, M., Schoot, B., van Heijenoort, J., and Gutmann, L. (2000). Novel mechanism of beta-lactam resistance due to bypass of dd-transpeptidation in *Enterococcus faecium*. *J Biol Chem*, 275(22) :16490–16496.

Mainardi, J.-L., Morel, V., Fourgeaud, M., Cremniter, J., Blanot, D., Legrand, R., Frehel, C., Arthur, M., Van Heijenoort, J., and Gutmann, L. (2002). Balance between two transpeptidation mechanisms determines the expression of β-lactam resistance in *Enterococcus faecium*. *J Biol Chem*, 277(39) :35801–35807.

Mainardi, J.-L., Villet, R., Bugg, T. D., Mayer, C., and Arthur, M. (2008). Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in gram-positive bacteria. *FEMS Microbiol Rev*, 32(2) :386–408.

Majiduddin, F. K., Materon, I. C., and Palzkill, T. G. (2002). Molecular analysis of beta-lactamase structure and function. *Int J Med Microbiol*, 292(2) :127–137.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems : enhanced expression and solubility with SUMO. *Protein Sci*, 15(1) :182–189.

Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and WÃ¼thrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. iupac-iubmb-iupab inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *J Biomol NMR*, 12(1) :1–23.

Marquardt, J. L., Brown, E. D., Lane, W. S., Haley, T. M., Ichikawa, Y., Wong, C. H., and Walsh, C. T. (1994). Kinetics, stoichiometry, and identification of the reactive thiolate in the inactivation of UDP-GlcNac enolpyruvoyl transferase by the antibiotic fosfomycin. *Biochemistry*, 33(35) :10646–10651.

Matias, V. R. F., Al-Amoudi, A., Dubochet, J., and Beveridge, T. J. (2003). Cryo-transmission electron microscopy of frozen-hydrated sections of *Escherichia coli* and *Pseudomonas aeruginosa*. *J Bacteriol*, 185(20) :6112–6118.

Matias, V. R. F. and Beveridge, T. J. (2005). Cryo-electron microscopy reveals native polymeric cell wall structure in *Bacillus subtilis* 168 and the existence of a periplasmic space. *Mol Microbiol*, 56(1) :240–251.

Matsuhashi, M., Wachi, M., and Ishino, F. (1990). Machinery for cell growth and division : penicillin-binding proteins and other proteins. *Res Microbiol*, 141(1) :89–103.

Maurer, T. (2005). NMR studies of protein-ligand interactions. *Methods Mol Biol*, 305 :197–214.

Mehra-Chaudhary, R., Mick, J., and Beamer, L. J. (2011). Crystal structure of bacillus anthracis phosphoglucosamine mutase, an enzyme in the peptidoglycan biosynthetic pathway. *J Bacteriol*, 193(16) :4081–4087.

Meiboom, S. and Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.*, 29 :688–691.

Mengin-Lecreulx, D., Texier, L., Rousseau, M., and van Heijenoort, J. (1991). The murG gene of *Escherichia coli* codes for the UDP-N-acetylglucosamine : N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *J Bacteriol*, 173(15) :4625–4636.

Mengin-Lecreulx, D. and van Heijenoort, J. (1994). Copurification of glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphate uridyltransferase activities of *Escherichia coli* : characterization of the glmU gene product as a bifunctional enzyme catalyzing two subsequent steps in the pathway for UDP-N-acetylglucosamine synthesis. *J Bacteriol*, 176(18) :5788–5795.

Meroueh, S. O., Bencze, K. Z., Hesek, D., Lee, M., Fisher, J. F., Stemmler, T. L., and Mobashery, S. (2006). Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci U S A*, 103(12) :4404–4409.

Michaud, C., Blanot, D., Flouret, B., and Van Heijenoort, J. (1987). Partial purification and specificity studies of the D-glutamate-adding and D-alanyl-D-alanine-adding enzymes from *Escherichia coli* k12. *Eur J Biochem*, 166(3) :631–637.

Miller, E. L. (2002). The penicillins : a review and update. *J Midwifery Womens Health*, 47(6) :426–434.

Millet, O., Muhandiram, D. R., Skrynnikov, N. R., and Kay, L. E. (2002). Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in ^{13}C-labeled and fractionally ^{2}H-enriched proteins in solution. *J Am Chem Soc*, 124(22) :6439–6448.

Mittermaier, A. and Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, 312(5771) :224–228.

Mohammadi, T., van Dam, V., Sijbrandi, R., Vernet, T., Zapun, A., Bouhss, A., Diepeveen-de Bruin, M., Nguyen-Distèche, M., de Kruijff, B., and Breukink, E. (2011). Identification of ftsw as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J*, 30(8) :1425–1432.

Morlot, C., Pernot, L., Le Gouellec, A., Di Guilmi, A. M., Vernet, T., Dideberg, O., and Dessen, A. (2005). Crystal structure of a peptidoglycan synthesis regulatory factor (PBP3) from *Streptococcus pneumoniae*. *J Biol Chem*, 280(16) :15984–15991.

Mukherjee, S., Pondaven, S. P., and Jaroniec, C. P. (2011). Conformational flexibility of a human immunoglobulin light chain variable domain by relaxation dispersion nuclear magnetic resonance spectroscopy : implications for protein misfolding and amyloid assembly. *Biochemistry*, 50(26) :5845–5857.

Murat, D., Byrne, M., and Komeili, A. (2010). Cell biology of prokaryotic organelles. *Cold Spring Harb Perspect Biol*, 2(10) :a000422.

Nakamura, H. (1996). Roles of electrostatic interaction in proteins. *Q Rev Biophys*, 29(1) :1–90.

Nanninga, N. (1998). Morphogenesis of *Escherichia coli*. *Microbiol Mol Biol Rev*, 62(1) :110–129.

Neuhaus, F. C. and Baddiley, J. (2003). A continuum of anionic charge : structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol Mol Biol Rev*, 67(4) :686–723.

Nicholas, R. A., Krings, S., Tomberg, J., Nicola, G., and Davies, C. (2003). Crystal structure of wild-type penicillin-binding protein 5 from *Escherichia coli* : implications for deacylation of the acyl-enzyme complex. *J Biol Chem*, 278(52) :52826–52833.

Nielsen, J. E., Gunner, M. R., and GarcÃa-Moreno, B. E. (2011). The pka cooperative : a collaborative effort to advance structure-based calculations of pka values and electrostatic effects in proteins. *Proteins*, 79(12) :3249–3259.

Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. automated assignment of ambiguous noe crosspeaks and disulphide connectivities. *J Mol Biol*, 245(5) :645–660.

Nilges, M., Macias, M. J., O'Donoghue, S. I., and Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints : the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol*, 269(3) :408–422.

Norrby, S. R., Alestig, K., Björnegard, B., Burman, L. A., Ferber, F., Huber, J. L., Jones, K. H., Kahan, F. M., Kahan, J. S., Kropp, H., Meisinger, M. A., and Sundelof, J. G. (1983). Urinary recovery of N-formimidoyl thienamycin (MK0787) as affected by coadministration of N-formimidoyl thienamycin dehydropeptidase inhibitors. *Antimicrob Agents Chemother*, 23(2) :300–307.

Norrby, S. R., Nord, C. E., and R., F. (2005). Lack of development of new antimicrobial drugs : a potential serious threat to public health. *Lancet Infect Dis.*, 2 :115–119.

Okuda, S. and Tokuda, H. (2011). Lipoprotein sorting in bacteria. *Annu Rev Microbiol*, 65 :239–259.

Olsen, L. R. and Roderick, S. L. (2001). Structure of the *Escherichia coli* GlmU pyrophosphorylase and acetyltransferase active sites. *Biochemistry*, 40(7) :1913–1921.

Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2009). Protein ionizable groups : pk values and their contribution to protein stability and solubility. *J Biol Chem*, 284(20) :13285–13289.

Palmer, 3rd, A. (2001). NMR probes of molecular dynamics : overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct*, 30 :129–155.

Palmer, 3rd, A., Rance, M., and Wright, P. E. (1991). ntramolecular motions of a zinc finger DNA-binding domain from xfin characterized by proton-detected natural abundance ^{13}C heteronuclear NMR spec- troscopy. *J Am Chem Soc*, 113 :4371–4380.

Palmer, 3rd, A. G. (2004). Nmr characterization of the dynamics of biomacromolecules. *Chem Rev*, 104(8) :3623–3640.

Panavas, T., Sanders, C., and Butt, T. R. (2009). SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods Mol Biol*, 497 :303–317.

Papp-Wallace, K. M., Endimiani, A., Taracila, M. A., and Bonomo, R. A. (2011). Carbapenems : past, present, and future. *Antimicrob Agents Chemother*, 55(11) :4943–4960.

Pares, S., Mouz, N., Pétillot, Y., Hakenbeck, R., and Dideberg, O. (1996). X-ray structure of *Streptococcus pneumoniae* PBP2x, a primary penicillin target enzyme. *Nat Struct Biol*, 3(3) :284–289.

Patti, G. J., Kim, S. J., and Schaefer, J. (2008). Characterization of the peptidoglycan of vancomycin-susceptible enterococcus faecium. *Biochemistry*, 47(32) :8378–8385.

Paul, K., Brunstetter, D., Titen, S., and Blair, D. F. (2011). A molecular mechanism of direction switching in the flagellar motor of *Escherichia coli*. *Proc Natl Acad Sci U S A*, 108(41) :17171–17176.

Paulsen, I. T., Brown, M. H., and Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. *Microbiol Rev*, 60(4) :575–608.

Pechère, J. C. (2001). New perspectives on macrolide antibiotics. *Int J Antimicrob Agents*, 18 :S93–S97.

Pellecchia, M., Fattorusso, R., and Wider, G. (1998). Determination of the dihedral angle ψ based on J coupling measurements in ^{15}N/^{13}C-labeled proteins. *J Am Chem Soc*, 120 :6824–6825.

Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993). Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated iiiglc, a signal-transducing protein from escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci*, 2(4) :543–558.

Peng, J. W. and Wagner, V. T. G. (1991). 2D heteronuclear NMR measurements of spin-lattice relaxation times in the rotating frame of x nuclei in heteronuclear HX spin systems. *J Magn Reson*, 94(1) :82–100.

Pervushin, K., Riek, R., Wider, G., and WÃ¼thrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A*, 94(23) :12366–12371.

Pink, D., Moeller, J., Quinn, B., Jericho, M., and Beveridge, T. (2000). On the architecture of the gram-negative bacterial murein sacculus. *J Bacteriol*, 182(20) :5925–5930.

Pisabarro, A. G., de Pedro, M. A., and VÃ¡zquez, D. (1985). Structural modifications in the peptidoglycan of *Escherichia coli* associated with changes in the state of growth of the culture. *J Bacteriol*, 161(1) :238–242.

Poole, K. (2004). Resistance to beta-lactam antibiotics. *Cell Mol Life Sci*, 61(17) :2200–2223.

Pootoolal, J., Neu, J., and Wright, G. D. (2002). Glycopeptide antibiotic resistance. *Annu Rev Pharmacol Toxicol*, 42 :381–408.

Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004). Residual dipolar couplings in structure determination of biomolecules. *Chem Rev*, 104(8) :3519–3540.

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol*, 7 :95–99.

Rasia, R. M., Lescop, E., Palatnik, J. F., Boisbouvier, J., and Brutscher, B. (2011). Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *J Biomol NMR*, 51(3) :369–378.

Rasmussen, J. R. and Strominger, J. L. (1978). Utilization of a depsipeptide substrate for trapping acyl-enzyme intermediates of penicillin-sensitive D-alanine carboxypeptidases. *Proc Natl Acad Sci U S A*, 75(1) :84–88.

Religa, T. L., Ruschak, A. M., Rosenzweig, R., and Kay, L. E. (2011). Site-directed methyl group labeling as an NMR probe of structure and dynamics in supramolecular protein systems : applications to the proteasome and to the ClpP protease. *J Am Chem Soc*, 133(23) :9063–9068.

Reynolds, W. F., Peat, I. R., Freedman, M. H., and Lyerla, Jr, J. (1973). Determination of the tautomeric form of the imidazole ring of L-histidine in basic solution by carbon-13 magnetic resonance spectroscopy. *J Am Chem Soc*, 95(2) :328–331.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2007). ARIA2 : automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3) :381–382.

Rogers, H., Perkins, H., and J.B., W. (1980). *Microbial cell walls and membranes*. Chapman and Hall, London, UK.

Rückert, M. and Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc*, 122(4) :7793–7797.

Ruschak, A. M. and Kay, L. E. (2010). Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR*, 46(1) :75–87.

Salton, M. R. J. and Horne, R. W. (1951). Studies of the bacterial cell wall. ii. methods of preparation and some properties of cell walls. *Biochim Biophys Acta*, 7(2) :177–197.

Sattler, M., Schleucherb, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, 34 :93–158.

Sauvage, E., Kerff, F., Fonzé, E., Herman, R., Schoot, B., Marquette, J. P., Taburet, Y., Prevost, D., Dumas, J., Leonard, G., Stefanic, P., Coyette, J., and Charlier, P. (2002). The 2.4-Å crystal structure of the penicillin-resistant penicillin-binding protein PBP5fm from *Enterococcus faecium* in complex with benzylpenicillin. *Cell Mol Life Sci*, 59(7) :1223–1232.

Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A., and Charlier, P. (2008). The penicillin-binding proteins : structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev*, 32(2) :234–258.

Schmieder, R. and Edwards, R. (2012). Insights into antibiotic resistance through metagenomic approaches. *Future Microbiol*, 7(1) :73–89.

Schumann, F. H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.-P., and Kalbitzer, H. R. (2007). Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J Biomol NMR*, 39(4) :275–289.

Schüttelkopf, A. W. and van Aalten, D. M. F. (2004). PRODRG : a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 8) :1355–1363.

Segur, J. B. and Oberstar, H. E. (1951). Viscosity of glycerol and its aqueous solutions. *Industrial and Engineering Chemistry*, 43(9) :2117–2120.

Sharif, S., Singh, M., Kim, S. J., and Schaefer, J. (2009). *Staphylococcus aureus* peptidoglycan tertiary structure from carbon-13 spin diffusion. *J Am Chem Soc*, 131(20) :7023–7030.

Shaw, K. J. and Barbachyn, M. R. (2011). The oxazolidinones : past, present, and future. *Ann N Y Acad Sci*, 1241 :48–70.

Sheehan, J. and Henery-Logan, K. (1959). A general synthesis of the penicillins. *J Am Chem Soc*, 81 :5838–5839.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). Talos+ : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*, 44(4) :213–223.

Shi, Q., Meroueh, S. O., Fisher, J. F., and Mobashery, S. (2011). A computational evaluation of the mechanism of penicillin-binding protein-catalyzed cross-linking of the bacterial cell wall. *J Am Chem Soc*, 133(14) :5274–5283.

Silvaggi, N. R., Josephine, H. R., Kuzin, A. P., Nagarajan, R., Pratt, R. F., and Kelly, J. A. (2005). Crystal structures of complexes between the R61 DD-peptidase and peptidoglycan-mimetic β-lactams : a non-covalent complex with a "perfect penicillin". *J Mol Biol*, 345(3) :521–533.

Skarzynski, T., Mistry, A., Wonacott, A., Hutchinson, S. E., Kelly, V. A., and Duncan, K. (1996). Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. *Structure*, 4(12) :1465–1474.

Skrynnikov, N. R., Millet, O., and Kay, L. E. (2002). Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions. *J Am Chem Soc*, 124(22) :6449–6460.

Slupsky, C. M., Gentile, L. N., and McIntosh, L. P. (1998). Assigning the NMR spectra of aromatic amino acids in proteins : analysis of two Ets pointed domains. *Biochem Cell Biol*, 76(2-3) :379–390.

Spízek, J., Novotná, J., and Rezanka, T. (2004). Lincosamides : chemical structure, biosynthesis, mechanism of action, resistance, and applications. *Adv Appl Microbiol*, 56 :121–154.

Spyracopoulos, L. and Sykes, B. D. (2001). Thermodynamic insights into proteins from NMR spin relaxation studies. *Curr Opin Struct Biol*, 11(5) :555–559.

Steenbergen, J. N., Alder, J., Thorne, G. M., and Tally, F. P. (2005). Daptomycin : a lipopeptide antibiotic for the treatment of serious gram-positive infections. *J Antimicrob Chemother*, 55(3) :283–288.

Stone, M. J. (2001). NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc Chem Res*, 34(5) :379–388.

Stowell, M. H., Miyazawa, A., and Unwin, N. (1998). Macromolecular structure determination by electron microscopy : new advances and recent results. *Curr Opin Struct Biol*, 8(5) :595–600.

Struve, W. G. and Neuhaus, F. C. (1965). Evidence for an initial acceptor of UDP-nac-muramyl-pentapeptide in the synthesis of bacterial mucopeptide. *Biochem Biophys Res Commun*, 18 :6–12.

Sudmeier, J. L., Bradshaw, E. M., Haddad, K. E. C., Day, R. M., Thalhauser, C. J., Bullock, P. A., and Bachovchin, W. W. (2003). Identification of histidine tautomers in proteins by 2d ^{1}H/^{13}C(delta2) one-bond correlated NMR. *J Am Chem Soc*, 125(28) :8430–8431.

Sung, M.-T., Lai, Y.-T., Huang, C.-Y., Chou, L.-Y., Shih, H.-W., Cheng, W.-C., Wong, C.-H., and Ma, C. (2009). Crystal structure of the membrane-bound bifunctional transglycosylase PBP1b from *Escherichia coli*. *Proc Natl Acad Sci U S A*, 106(22) :8824–8829.

Tenson, T., Lovmar, M., and Ehrenberg, M. (2003). The mechanism of action of macrolides, lincosamides and streptogramin b reveals the nascent peptide exit path in the ribosome. *J Mol Biol*, 330(5) :1005–1014.

Teplyakov, A., Obmolova, G., Badet, B., and Badet-Denisot, M. A. (2001). Channeling of ammonia in glucosamine-6-phosphate synthase. *J Mol Biol*, 313(5) :1093–1102.

Thaker, M., Spanogiannopoulos, P., and Wright, G. (2010). The tetracycline resistome. *Cell*, 67 :419–431.

Tian, F., Valafar, H., and Prestegard, J. H. (2001). A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. *J Am Chem Soc*, 123(47) :11791–11796.

Tipper, D. J. and Strominger, J. L. (1965). Mechanism of action of penicillins : a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci U S A*, 54(4) :1133–1141.

Tjandra, N. and Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 278(5340) :1111–1114.

Tjandra, N., Kuboniwa, H., Ren, H., and Bax, A. (1995). Rotational dynamics of calcium-free calmodulin studied by ^{15}N-NMR relaxation measurements. *Eur J Biochem*, 230(3) :1014–1024.

Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins : information for structure determination in solution. *Proc Natl Acad Sci U S A*, 92(20) :9279–9283.

Touhami, A., Jericho, M. H., and Beveridge, T. J. (2004). Atomic force microscopy of cell growth and division in staphylococcus aureus. *J Bacteriol*, 186(11) :3286–3295.

Triboulet, S., Arthur, M., Mainardi, J.-L., Veckerlé, C., Dubée, V., Nguekam-Moumi, A., Gutmann, L., Rice, L. B., and Hugonnet, J.-E. (2011). Inactivation kinetics of a new target of beta-lactam antibiotics. *J Biol Chem*, 286(26) :22777–22784.

Tugarinov, V., Kanelis, V., and Kay, L. E. (2006). Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc*, 1(2) :749–754.

Tugarinov, V. and Kay, L. E. (2003). Ile, leu, and val methyl assignments of the 723-residue malate synthase g using a new labeling strategy and novel NMR methods. *J Am Chem Soc*, 125(45) :13868–13878.

Tycko, R., Blanco, F. J., and Ishii, Y. (2000). Alignment of biopolymers in strained gels a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J Am Chem Soc*, 122(38) :9340–9341.

Typas, A., Banzhaf, M., Gross, C. A., and Vollmer, W. (2012). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol*, 10(2) :123–136.

van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des*, 10(3) :255–262.

van Bambeke, F., Balzi, E., and Tulkens, P. M. (2000). Antibiotic efflux pumps. *Biochem Pharmacol*, 60(4) :457–470.

van Heijenoort, J. (2001a). Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology*, 11(3) :25R–36R.

van Heijenoort, J. (2001b). Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat Prod Rep*, 18(5) :503–519.

van Heijenoort, Y., Gòmez, M., Derrien, M., Ayala, J., and van Heijenoort, J. (1992). Membrane intermediates in the peptidoglycan metabolism of escherichia coli : possible roles of PBP 1b and PBP 3. *J Bacteriol*, 174(11) :3549–3557.

Van Melckebeke, H., Simorre, J.-P., and Brutscher, B. (2004). Amino acid-type edited NMR experiments for methyl-methyl distance measurement in ^{13}C-labeled proteins. *J Am Chem Soc*, 126(31) :9584–9591.

Vila, J. A., Arnautova, Y. A., Vorobjev, Y., and Scheraga, H. A. (2011). Assessing the fractions of tautomeric forms of the imidazole ring of histidine in proteins as a function of ph. *Proc Natl Acad Sci U S A*, 108(14) :5602–5607.

Voet, D. and Voet, J. G. (2004). *Book : Biochemistry 3rd edition*. J. Wiley & Sons.

Vogt, A. D. and Di Cera, E. (2012). Conformational selection or induced fit ? a critical appraisal of the kinetic mechanism. *Biochemistry*, 51 :5894–5902.

Vollmer, W. (2006). The prokaryotic cytoskeleton : a putative target for inhibitors and antibiotics ? *Appl Microbiol Biotechnol*, 73(1) :37–47.

Vollmer, W., Blanot, D., and de Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiol Rev*, 32(2) :149–167.

Vollmer, W. and Höltje, J.-V. (2004). The architecture of the murein (peptidoglycan) in gram-negative bacteria : vertical scaffold or horizontal layer(s) ? *J Bacteriol*, 186(18) :5978–5987.

Vollmer, W. and Seligman, S. J. (2010). Architecture of peptidoglycan : more data and more models. *Trends Microbiol*, 18(2) :59–66.

Vuister, G. and Bax, A. (1993). Quantitative J correlation : A new approach for measuring homonuclear three-bond $J(H^N H^\alpha)$ coupling constants in ^{15}N-enriched proteins. *J Am Chem Soc*, 115 :7772–7777.

Wagner, G., Hyberts, S. G., and Havel, T. F. (1992). NMR structure determination in solution : a critique and comparison with X-ray crystallography. *Annu Rev Biophys Biomol Struct*, 21 :167–198.

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT : a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, 8(2) :127–134.

Walsh, C. (2003). Where will new antibiotics come from ? *Nat Rev Microbiol*, 1(1) :65–70.

Wand, A. J. and Englander, S. W. (1996). Protein complexes studied by NMR spectroscopy. *Curr Opin Biotechnol*, 7(4) :403–408.

Wang, Z., Fast, W., Valentine, A. M., and Benkovic, S. J. (1999). Metallo-beta-lactamase : structure and mechanism. *Curr Opin Chem Biol*, 3(5) :614–622.

Ward, J. B. and Perkins, H. R. (1973). The direction of glycan synthesis in a bacterial peptidoglycan. *Biochem J*, 135(4) :721–728.

Webber, M. A. and Piddock, L. J. V. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother*, 51(1) :9–11.

Weidenmaier, C. and Peschel, A. (2008). Teichoic acids and related cell-wall glycopolymers in gram-positive physiology and host interactions. *Nat Rev Microbiol*, 6(4) :276–287.

Weikl, T. R. and von Deuster, C. (2009). Selected-fit versus induced-fit protein binding : kinetic differences and mutational analysis. *Proteins*, 75(1) :104–110.

Welzel, P. (2005). Syntheses around the transglycosylation step in peptidoglycan biosynthesis. *Chem Rev*, 105(12) :4610–4660.

Weston, A. and Perkins, H. R. (1977). Biosynthesis of wall-linked teichuronic acid by a wall-plus-membrane preparation from micrococcus luteus. effect of antibiotics. *FEBS Lett*, 76(2) :195–198.

Wheeler, R., Mesnage, S., Boneca, I. G., Hobbs, J. K., and Foster, S. J. (2011). Super-resolution microscopy reveals cell wall dynamics and peptidoglycan architecture in ovococcal bacteria. *Mol Microbiol*, 82(5) :1096–1109.

White, C. L., Kitich, A., and Gober, J. W. (2010). Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins mreb and mred. *Mol Microbiol*, 76(3) :616–633.

WHO (2011). *Global tuberculosis control 2010*. World Health Organization report 2011. http ://www.who.int/tb/en/br.

Wilke, M. S., Lovering, A. L., and Strynadka, N. C. J. (2005). Beta-lactam antibiotic resistance : a current structural perspective. *Curr Opin Microbiol*, 8(5) :525–533.

Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985). Solution conformation of proteinase inhibitor iia from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J Mol Biol*, 182(2) :295–315.

Wishart, D. S. and Sykes, B. D. (1994). The ^{13}C chemical-shift index : a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR*, 4(2) :171–180.

Wright, A. J. (1999). The penicillins. *Mayo Clin Proc*, 74(3) :290–307.

Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993). Two-dimensional NMR experiments for correlating $^{13}C\beta$ and $^{1}H\delta/\epsilon$ chemical shifts of aromatic residues in ^{13}C-labeled proteins via scalar couplings. *J. Am. Chem. Soc.*, 115 :11054–11055.

Yao, X., Jericho, M., Pink, D., and Beveridge, T. (1999). Thickness and elasticity of gram-negative murein sacculi measured by atomic force microscopy. *J Bacteriol*, 181(22) :6865–6875.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C., and Shively, J. M. (2008). Protein-based organelles in bacteria : carboxysomes and related microcompartments. *Nat Rev Microbiol*, 6(9) :681–691.

Yousif, S. Y., Broome-Smith, J. K., and Spratt, B. G. (1985). Lysis of *Escherichia coli* by beta-lactam antibiotics : deletion analysis of the role of penicillin-binding proteins 1A and 1B. *J Gen Microbiol*, 131(10) :2839–2845.

Yuan, Y., Barrett, D., Zhang, Y., Kahne, D., Sliz, P., and Walker, S. (2007). Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Proc Natl Acad Sci U S A*, 104(13) :5348–5353.

Zapun, A., Contreras-Martel, C., and Vernet, T. (2008). Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol Rev*, 32(2) :361–385.

Zhu, G., Xia, Y., Lin, D., and Gao, X. (2004). Trosy-based correlation and noe spectroscopy for NMR structural studies of large proteins. *Methods Mol Biol*, 278 :57–78.

Zhuang, T., Leffler, H., and Prestegard, J. H. (2006). Enhancement of bound-state residual dipolar couplings : conformational analysis of lactose bound to Galectin-3. *Protein Sci*, 15(7) :1780–1790.

Zorzi, W., Zhou, X. Y., Dardenne, O., Lamotte, J., Raze, D., Pierre, J., Gutmann, L., and Coyette, J. (1996). Structure of the low-affinity penicillin-binding protein 5 PBP5fm in wild-type and highly penicillin-resistant strains of *Enterococcus faecium*. *J Bacteriol*, 178(16) :4948–4957.

Zuiderweg, E. R. P. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry*, 41(1) :1–7.

Annexes

Annexe 1: Amino acids nomenclature in NMR

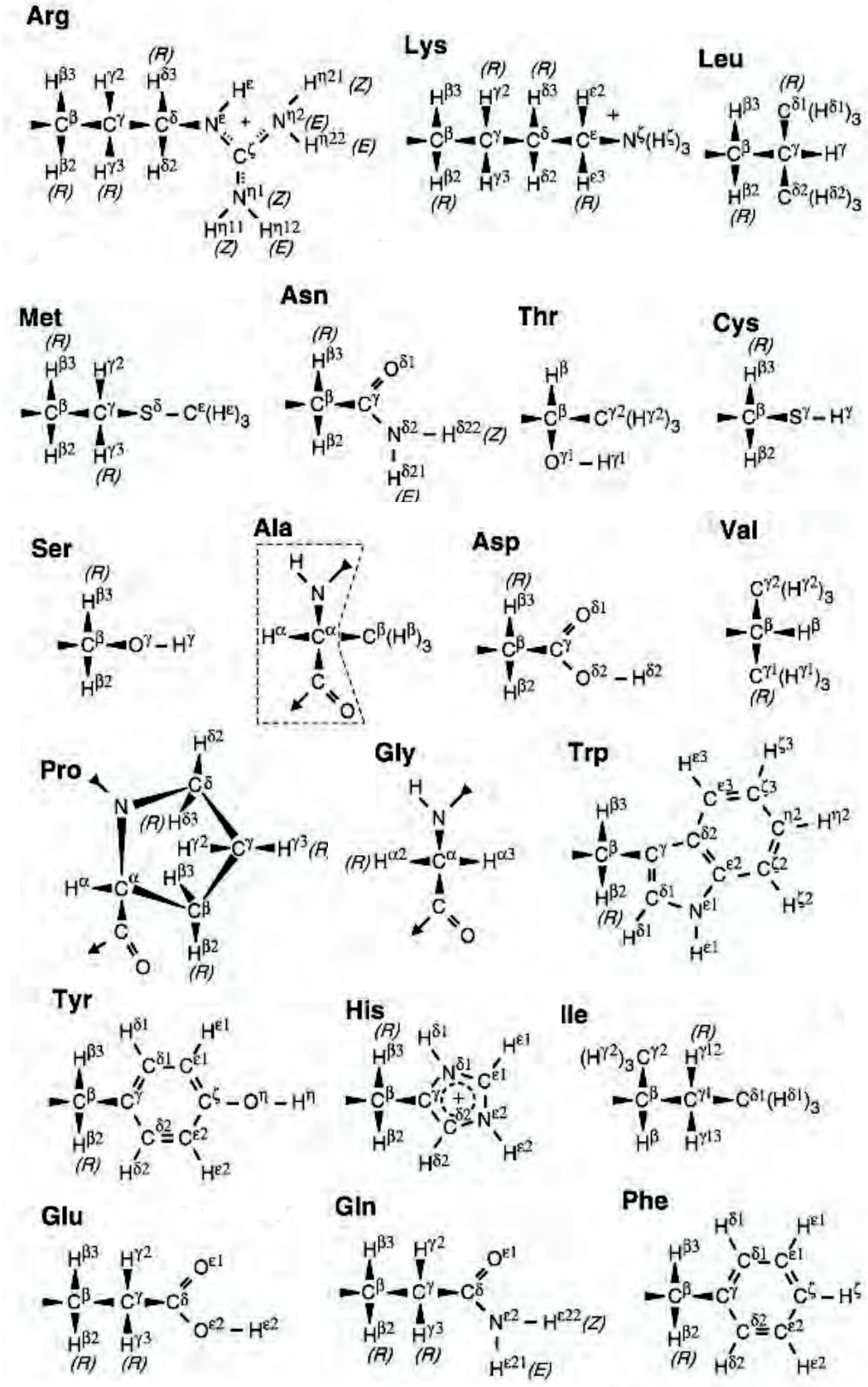


Figure 7.23: Amino acids nomenclature following IUPAC rules. Taken from (Markley *et al.*, 1998).

Annexe 2: Modified files in ARIA distribution

Annexe 2 contains the seven files that must be modified in the ARIA distribution when defining a new residue. For structure calculations of LDts with a covalent bond between the sulfur atom of the catalytic cysteine and the carbonyl of the β-lactam cycle of carbapenems, it was required to create a new residue in ARIA. I will therefore give example of the modified python and CNS files for the cysteine bound to imipenem antibiotic, which is referred to as "IMP" in the mentioned files. Figure 7.24 displays the structure of IMP residue with the corresponding atom names used in ARIA files.

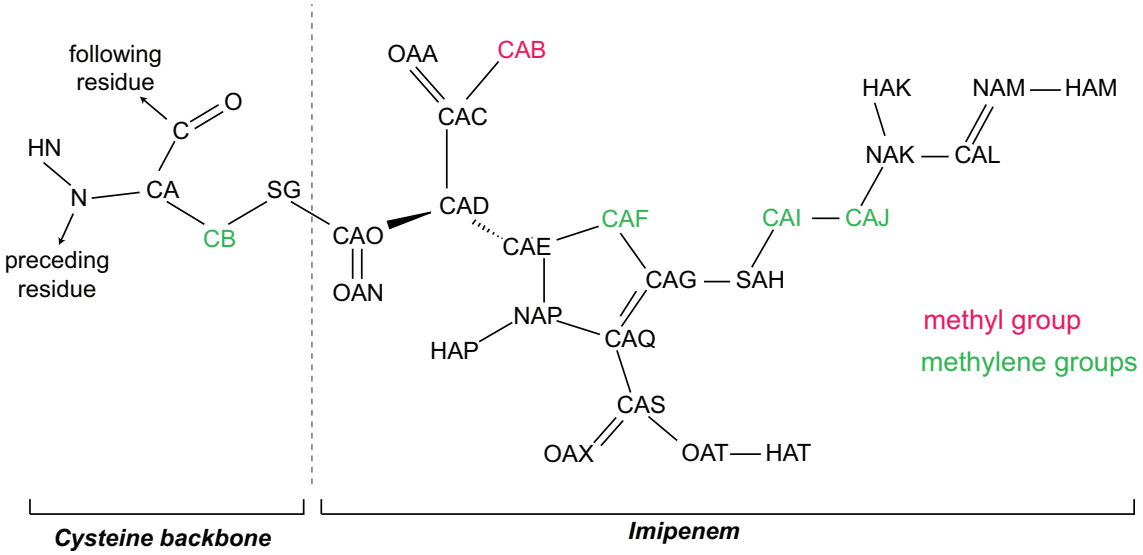


Figure 7.24: 2D structure of the modified cysteine defined in ARIA to integrate the imipenem-bound antibiotic.

The following files must be modified within ARIA:

- **Annexe 2.1: AminoAcid.py** in the directory src/py/legacy of the ARIA distribution, to define one-letter and three-letters codes for the new residue.
- **Annexe 2.2: PseudoAtom.py** in the directory src/py/legacy of the ARIA distribution, to define methylene and methyl groups.
- **Annexe 2.3: Nomenclature.py** in the directory src/py/legacy of the ARIA distribution, to define methylenes and methyls.
- **Annexe 2.4: iupac.xml** in the directory src/py/data of the ARIA distribution. In this file must be added all the atom names and atom types of the new residue. Backbone, sidechain, methylenes, methyls and terminus atoms must also be declared.
- **Annexe 2.5: atomnames.xml** in the directory src/py/data of the ARIA distribution. This file contains the atom names for every residues, with IUPAC, CNS and DYANA nomenclatures.
- **Annexe 2.6: topalldg5.3.pro** in the aria2/cns/toppar directory. In this file must be copied the CNS topologies generated by PRODRG. It defines all the existing atoms with their atom type and their charges, and defines which atoms are linked to other atoms through covalent bond, angles, and impropers.
- **Annexe 2.7: paralldg5.3.pro** in the aria2/cns/toppar directory. In this file must be copied the CNS parameters generated by PRODRG. Only the parameters involving atoms with new atom

type must be added. This file contains all the bond, angle, and improper values for every linked atom-types.

Only sparse selections of these files are shown in the Annexes. The modifications are highlighted in yellow. For the two last files, the modifications requires the previous determination of molecular topologies using PRODRG program as explained in section 3.2.3. PRODRG can provide the molecular topologies and parameters in the CNS file format. These two files must be copied within the topalldg5.3.pro and the paralldg5.3.pro files, respectively.

For the atoms encountering similar neighbors and bindings that already exist in the 20 common amino acids defined in ARIA, the atom type might be replaced by the existing atom type in topallhdg5.3.pro. paralldg5.3.pro can then be simplified by removing all the definitions that are duplicated. When the atoms are completely new in terms of structure, a new type must be defined. For IMP residue, the atoms corresponding to the cysteine backbone, as well as all the atoms located in linear aliphatic chains, were declared as belonging to an existing atom type. As an example, aliphatic protons are defined as “H” type, methyl carbons as “CH3E”, methylene carbons as “CH2E”, etc... For aromatics and cycles, it is more complicated as the angles and bond lenghts differs than in linear aliphatic chains. New atom type was therefore necessary for the following atoms: CAE, CAF, NAP, HAP, CAQ, CAS, OAX, OAT, CAG, CAL, NAM and HAM.

Annexe 2.1: AminoAcid.py

```
"""
A module to deal with aminoacids in 3- or 1-letter code

def AminoAcid(input):
    """
    INPUT:  a string containing an amino acid in 3- or 1-letter code
            (only the 20 common amino acids can be converted)
            upper- or lowercase; spaces, tabs and linebreaks are neglected
    OUTPUT: returns a list of strings with the 1-letter-code as element 0,
            the 3-letter-code as element 1 and the name as element 2.
               all returned 1- and 3-letter codes are uppercase, the name
            is lowercase.
    if the input is not a common amino acid, the elements 0 and 1 will be
    returned empty with the input string as third element.
    """
    import string
    one2all ={'A': ('A', 'ALA', 'alanine'),
              'R': ('R', 'ARG', 'arginine'),
              'N': ('N', 'ASN', 'asparagine'),
              'D': ('D', 'ASP', 'aspartic acid'),
              'C': ('C', 'CYS', 'cysteine'),
              'Q': ('Q', 'GLN', 'glutamine'),
              'E': ('E', 'GLU', 'glutamic acid'),
              'G': ('G', 'GLY', 'glycine'),
              'H': ('H', 'HIS', 'histidine'),
              'I': ('I', 'ILE', 'isoleucine'),
              'L': ('L', 'LEU', 'leucine'),
              'K': ('K', 'LYS', 'lysine'),
              'M': ('M', 'MET', 'methionine'),
              'F': ('F', 'PHE', 'phenylalanine'),
              'P': ('P', 'PRO', 'proline'),
              'S': ('S', 'SER', 'serine'),
              'T': ('T', 'THR', 'threonine'),
              'W': ('W', 'TRP', 'tryptophan'),
              'Y': ('Y', 'TYR', 'tyrosine'),
              'V': ('V', 'VAL', 'valine'),
              'Z': ('Z', 'IMP', 'cystein_imp'),}
    three2all = {'ALA': ('A', 'ALA', 'alanine'),
                 'ARG': ('R', 'ARG', 'arginine'),
                 'ASN': ('N', 'ASN', 'asparagine'),
                 'ASP': ('D', 'ASP', 'aspartic acid'),
                       ...
                 'CYS': ('C', 'CYS', 'cysteine'),
                 'VAL': ('V', 'VAL', 'Valine'),
                 'IMP': ('Z', 'IMP', 'cystein_imp'),}
...
```

Annexe 2.2: PseudoAtom.py

```
"""
This module provides methods for the conversion of cns atomnames in
pseudoatomes and vice versa.

The nomenclature of the pseudoatoms is in the same spirit as in:
Markley et al. JMB 280, 933-952 (1998)

Methods:
   Pseudo2Atom(atom)   e.g. 'QA' -> 'ha#'
                        'QR' -> 'hd# or name he# or name hz'
   Pseudo2Tuple(atom)  e.g. 'QA' -> ('ha#',)
                        'QR' -> ('hd#', 'he#', 'hz#')
   Atom2Pseudo(atomname, aminoacid)
                   e.g. 'HB1', 'TYR' -> ('QB', )
                        'HG11', 'VAL' ->  ('QG1', 'QQG')
   Pseudo2IupacTuple(threelettercode, atomname)


################################################################################
def Pseudo2Atom(atom):
   """
   converts pseudoatoms into a CNS readable format
   input:   atom name (can be a pseudoatom)
   output:  CNS readable atom name
   """
   replacedic = {'QA': 'ha#',
                 'QB': 'hb#',
                 'QAN': 'han#',
                 'QG': 'hg#',
                 'QG1': 'hg1#',
                 'QG2': 'hg2#',
                 'QQG': 'hg#',
                 'QD': 'hd#',
                 'QD1': 'hd1#',
                 'QD2': 'hd2#',
                 'QQD': 'hd#',
                 'QE': 'he#',
                 'QE2': 'he2#',
                 'QR': 'hd# or name he# or name hz',
                 'QZ': 'hz#',
                 'QH1': 'hh1#',
                 'QH2': 'hh2#',
                 'QAB': 'hab#',
                 'QAF': 'haf#',
                 'QAI': 'hai#',
                 'QAJ': 'haj#',}
   atom = string.upper(atom)
   atom = string.strip(atom)
   if replacedic.has_key(atom):
      atom = replacedic[atom]
   return atom
```

Annexe 2.2 (continued)

```
def Pseudo2IupacTuple(threelettercode, atomname):
    """
    converting pseudoatoms into tuples of atomnames in IUPAC nomenclature
    """

    threelettercode = string.strip(threelettercode)

    replacedic = {'ALA': {'QB':  ('HB1', 'HB2', 'HB3'),\
                          'HB%': ('HB1', 'HB2', 'HB3'),\
                          'HB':  ('HB1', 'HB2', 'HB3')},\
                  'IMP': {'QB':  ('HB2', 'HB3'),\
                          'HB%': ('HB2', 'HB3'),\
                          'HB':  ('HB2', 'HB3'),\
                          'QAB':  ('HAB1', 'HAB2', 'HAB3'),\
                          'HAB%': ('HAB1', 'HAB2', 'HAB3'),\
                          'HAB':  ('HAB1', 'HAB2', 'HAB3'),\
                          'QAF':  ('HAF2', 'HAF3'),\
                          'HAF%': ('HAF2', 'HAF3'),\
                          'HAF':  ('HAF2', 'HAF3'),\
                          'QAI':  ('HAI2', 'HAI3'),\
                          'HAI%': ('HAI2', 'HAI3'),\
                          'HAI':  ('HAI2', 'HAI3'),\
                          'QAJ':  ('HAJ2', 'HAJ3'),\
                          'HAJ%': ('HAJ2', 'HAJ3'),\
                          'HAJ':  ('HAJ2', 'HAJ3')},\
(...) }

################################################################################
def Atom2Pseudo(atomname, aminoacid):
    """
    INPUT:   cns atomname and aminoacid type in 3-letter code
    OUTPUT:  returns a tuple of possible pseudoatoms, best hit comes first
             within the tuple
             if it is not possible to find a pseudoatom, an empty tuple
             will be returned
    """
    pseudoatom = ()
    aminoacid = string.upper(aminoacid)
    aminoacid = string.strip(aminoacid)
    atomname = string.upper(atomname)
    atomname = string.strip(atomname)
    if aminoacid == 'ALA':
        if atomname in ['HB1', 'HB2', 'HB3', 'HB%', 'HB#']:
            pseudoatom = ('QB', )
    elif aminoacid == 'IMP':
        if atomname in ['HB1', 'HB2', 'HB3', 'HB%', 'HB#']:
            pseudoatom = ('QB', )
        elif atomname in ['HAB1', 'HAB2', 'HAB3', 'HAB%', 'HAB#']:
            pseudoatom = ('QAB', 'QQAB')
        elif atomname in ['HAF1', 'HAF2', 'HAF3', 'HAF%', 'HAF#']:
            pseudoatom = ('QAF', )
        elif atomname in ['HAI1', 'HAI2', 'HAI3', 'HAI%', 'HAI#']:
            pseudoatom = ('QAI', )
        elif atomname in ['HAJ1', 'HAJ2', 'HAJ3', 'HAJ%', 'HAJ#']:
            pseudoatom = ('QAJ', )
(...)
```

Annexe 2.3: Nomenclature.py

```
"""A module to convert atomnames to IUPAC nomenclature"""

import re, sys, os, string
import AminoAcid


def ConvertCnsProtonNames(residueName, atomName):
    """
    convert an atomname from XPLOR/CNS to IUPAC nomenclature (or vice versa)
    residueName: a string which contains 1- or 3-letter code, e.g. 'A' or 'ALA'
                 only the 20 common aminoacids are supported!
    atomName:    a string containing the atomnames, e.g. 'HG12'
    """
    #I. get a clean three-letter code and strip & uppercase the atomName
    threeLetter = AminoAcid.AminoAcid(residueName)[1]
    if threeLetter[2] == '':
        print 'WARNING: residue name', residueName, 'not understood'
        return atomName
    atomName = string.upper(string.strip(atomName))

    #II. methylenes
    #1. GLY HA:
    if threeLetter == 'GLY' and atomName == 'HA1':
        atomName = 'HA2'
    elif threeLetter == 'GLY' and atomName == 'HA2':
        atomName = 'HA1'

    #2. ARG, ASN, ASP, CYS, IMP, GLN, GLU, HIS, LEU, LYS, MET, PHE, PRO, SER, TRP, TYR HB%:
    elif threeLetter in ('ARG', 'ASN', 'ASP', 'CYS', 'IMP', 'GLN', 'GLU', 'HIS', 'LEU', 'LYS',\
                         'MET', 'PHE', 'PRO', 'SER', 'TRP', 'TYR') and \
                         atomName == 'HB3': atomName = 'HB1'
    elif threeLetter in ('ARG', 'ASN', 'ASP', 'CYS', 'IMP','GLN', 'GLU', 'HIS', 'LEU', 'LYS',\
                         'MET', 'PHE', 'PRO', 'SER', 'TRP', 'TYR') and \
                         atomName == 'HB1': atomName = 'HB3'
  (...)
    #7. IMP HAF:
    elif threeLetter == 'IMP' and atomName == 'HAF3':  atomName = 'HAF1'
    elif threeLetter == 'IMP' and atomName == 'HAF1':  atomName = 'HAF3'
    #8. IMP HAI:
    elif threeLetter == 'IMP' and atomName == 'HAI3':  atomName = 'HAI1'
    elif threeLetter == 'IMP' and atomName == 'HAI1':  atomName = 'HAI3'
    #9. IMP HAJ:
    elif threeLetter == 'IMP' and atomName == 'HAJ3': atomName = 'HAJ1'
    elif threeLetter == 'IMP' and atomName == 'HAJ1': atomName = 'HAJ3'

    #III. methyls:
    #1. ALA beta:
    elif threeLetter == 'ALA' and atomName == 'HB2': atomName = 'HB1'
    elif threeLetter == 'ALA' and atomName == 'HB1': atomName = 'HB2'
  (...)
    #8. IMP/ERT HAB:
    elif threeLetter == 'IMP' and atomName == 'HAB1': atomName = 'HAB2'
    elif threeLetter == 'IMP' and atomName == 'HAB2': atomName = 'HAB1'

    return atomName

(...)
```

Annexe 2.4: iupac.xml

```
(...)
<!DOCTYPE topology SYSTEM "topology1.0.dtd">
<topology description="Amino acid definitions following the IUPAC convention" name="IUPAC">
  <residue residue_type="AMINO_ACID" name="IMP">
    <atom name="C" atom_type="C" hetero="None"/>
    <atom name="H" atom_type="H" hetero="N"/>
    <atom name="CA" atom_type="C" hetero="None"/>
    <atom name="O" atom_type="O" hetero="None"/>
    <atom name="N" atom_type="N" hetero="None"/>
    <atom name="CB" atom_type="C" hetero="None"/>
    <atom name="HA" atom_type="H" hetero="CA"/>
    <atom name="O'" atom_type="O" hetero="None"/>
    <atom name="O''" atom_type="O" hetero="None"/>
    <atom name="NT" atom_type="N" hetero="None"/>
    <atom name="HT" atom_type="H" hetero="NT"/>
    <atom name="H1" atom_type="H" hetero="N"/>
    <atom name="H2" atom_type="H" hetero="N"/>
    <atom name="H3" atom_type="H" hetero="N"/>
    <atom name="SG" atom_type="S" hetero="None"/>
    <atom name="HB3" atom_type="H" hetero="CB"/>
    <atom name="HB2" atom_type="H" hetero="CB"/>
    <atom name="CAO" atom_type="C" hetero="None"/>
    <atom name="OAN" atom_type="O" hetero="None"/>
    <atom name="CAD" atom_type="C" hetero="None"/>
    <atom name="HAD" atom_type="H" hetero="CAD"/>
    <atom name="CAC" atom_type="C" hetero="None"/>
    <atom name="HAC" atom_type="H" hetero="CAC"/>
    <atom name="CAB" atom_type="C" hetero="None"/>
    <atom name="HAB1" atom_type="H" hetero="CAB"/>
    <atom name="HAB2" atom_type="H" hetero="CAB"/>
    <atom name="HAB3" atom_type="H" hetero="CAB"/>
    <atom name="OAA" atom_type="O" hetero="None"/>
    <atom name="HAA" atom_type="H" hetero="OAA"/>
    <atom name="CAE" atom_type="C" hetero="None"/>
    <atom name="HAE" atom_type="H" hetero="CAE"/>
    <atom name="CAF" atom_type="C" hetero="None"/>
    <atom name="HAF2" atom_type="H" hetero="CAF"/>
    <atom name="HAF3" atom_type="H" hetero="CAF"/>
    <atom name="NAP" atom_type="N" hetero="None"/>
    <atom name="HAP" atom_type="H" hetero="NAP"/>
    <atom name="CAQ" atom_type="C" hetero="None"/>
    <atom name="CAS" atom_type="C" hetero="None"/>
    <atom name="OAX" atom_type="O" hetero="None"/>
    <atom name="OAT" atom_type="O" hetero="None"/>
    <atom name="HAT" atom_type="H" hetero="OAT"/>
    <atom name="CAG" atom_type="C" hetero="None"/>
    <atom name="SAH" atom_type="S" hetero="None"/>
    <atom name="CAI" atom_type="C" hetero="None"/>
    <atom name="HAI2" atom_type="H" hetero="CAI"/>
    <atom name="HAI3" atom_type="H" hetero="CAI"/>
    <atom name="CAJ" atom_type="C" hetero="None"/>
    <atom name="HAJ2" atom_type="H" hetero="CAJ"/>
    <atom name="HAJ3" atom_type="H" hetero="CAJ"/>
    <atom name="NAK" atom_type="N" hetero="None"/>
    <atom name="HAK" atom_type="H" hetero="NAK"/>
    <atom name="CAL" atom_type="C" hetero="None"/>
    <atom name="HAL" atom_type="H" hetero="CAL"/>
    <atom name="NAM" atom_type="N" hetero="None"/>
    <atom name="HAM" atom_type="H" hetero="NAM"/>
    <backbone>
      <atom_name name="N"/>
      <atom_name name="CA"/>
      <atom_name name="HA"/>
      <atom_name name="C"/>
    </backbone>
    <sidechain>
      <atom_name name="CB"/>
      <atom_name name="HB2"/>
      <atom_name name="HB3"/>
      <atom_name name="SG"/>
      <atom_name name="CAO"/>
      <atom_name name="OAN"/>
      <atom_name name="CAD"/>
      <atom_name name="HAD"/>
      <atom_name name="CAC"/>
      <atom_name name="HAC"/>
      <atom_name name="CAB"/>
      <atom_name name="HAB1"/>
      <atom_name name="HAB2"/>
      <atom_name name="HAB3"/>
      <atom_name name="OAA"/>
      <atom_name name="HAA"/>
      <atom_name name="CAE"/>
      <atom_name name="HAE"/>
      <atom_name name="CAF"/>
      <atom_name name="HAF2"/>
      <atom_name name="HAF3"/>
```

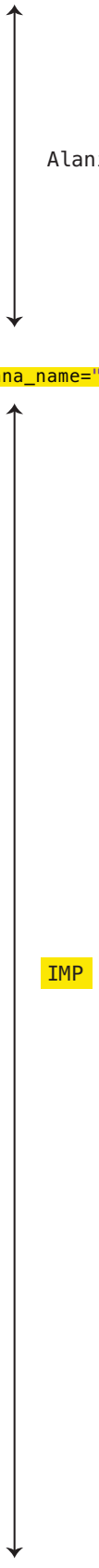
(continued)

```
      <atom_name name="NAP"/>
      <atom_name name="HAP"/>
      <atom_name name="CAQ"/>
      <atom_name name="CAS"/>
      <atom_name name="OAX"/>
      <atom_name name="OAT"/>
      <atom_name name="HAT"/>
      <atom_name name="CAG"/>
      <atom_name name="SAH"/>
      <atom_name name="CAI"/>
      <atom_name name="HAI2"/>
      <atom_name name="HAI3"/>
      <atom_name name="CAJ"/>
      <atom_name name="HAJ2"/>
      <atom_name name="HAJ3"/>
      <atom_name name="NAK"/>
      <atom_name name="HAK"/>
      <atom_name name="CAL"/>
      <atom_name name="HAL"/>
      <atom_name name="NAM"/>
      <atom_name name="HAM"/>
    </sidechain>
    <terminus terminus_type="N_STANDARD">
      <atom_name name="H"/>
    </terminus>
    <terminus terminus_type="N_AMINYL">
      <atom_name name="H1"/>
      <atom_name name="H2"/>
      <atom_name name="H3"/>
    </terminus>
    <terminus terminus_type="C_STANDARD">
      <atom_name name="O"/>
    </terminus>
    <terminus terminus_type="C_AMIDO">
      <atom_name name="O'"/>
      <atom_name name="NT"/>
      <atom_name name="HT"/>
    </terminus>
    <terminus terminus_type="C_CARBOXYL">
      <atom_name name="O'"/>
      <atom_name name="O''"/>
    </terminus>
    <equivalent_group group_type="METHYLENE">
      <atom_name name="HB2"/>
      <atom_name name="HB3"/>
    </equivalent_group>
    <equivalent_group group_type="METHYLENE">
      <atom_name name="HAF2"/>
      <atom_name name="HAF3"/>
    </equivalent_group>
    <equivalent_group group_type="METHYLENE">
      <atom_name name="HAI2"/>
      <atom_name name="HAI3"/>
    </equivalent_group>
    <equivalent_group group_type="METHYLENE">
      <atom_name name="HAJ2"/>
      <atom_name name="HAJ3"/>
    </equivalent_group>
    <equivalent_group group_type="METHYL">
      <atom_name name="HAB1"/>
      <atom_name name="HAB2"/>
      <atom_name name="HAB3"/>
    </equivalent_group>
    <equivalent_group group_type="NTERMINUS">
      <atom_name name="H1"/>
      <atom_name name="H2"/>
      <atom_name name="H3"/>
    </equivalent_group>
  </residue>
</topology>

(...)
```

Annexe 2.5: atomnames.xml

```
<conversion_table>
  <residue residue_type="AMINO_ACID" iupac_name="ALA" cns_name="ALA" dyana_name="ALA">
    <atom iupac_name="C" cns_name="C" dyana_name="C"/>
    <atom iupac_name="CA" cns_name="CA" dyana_name="CA"/>
    <atom iupac_name="CB" cns_name="CB" dyana_name="CB"/>
    <atom iupac_name="H" cns_name="HN" dyana_name="HN"/>
    <atom iupac_name="H1" cns_name="HT3" dyana_name="H1"/>
    <atom iupac_name="H2" cns_name="HT2" dyana_name="H2"/>
    <atom iupac_name="H3" cns_name="HT1" dyana_name="H3"/>
    <atom iupac_name="HA" cns_name="HA" dyana_name="HA"/>
    <atom iupac_name="HB1" cns_name="HB2" dyana_name="HB1"/>
    <atom iupac_name="HB2" cns_name="HB1" dyana_name="HB2"/>
    <atom iupac_name="HB3" cns_name="HB3" dyana_name="HB3"/>
    <atom iupac_name="HT" cns_name="None" dyana_name="None"/>
    <atom iupac_name="N" cns_name="N" dyana_name="N"/>
    <atom iupac_name="NT" cns_name="None" dyana_name="None"/>
    <atom iupac_name="O" cns_name="O" dyana_name="O"/>
    <atom iupac_name="O'" cns_name="OT1" dyana_name="O"/>
    <atom iupac_name="O''" cns_name="OT2" dyana_name="OXT"/>
  </residue>
  <residue residue_type="AMINO_ACID" iupac_name="IMP" cns_name="IMP" dyana_name="IMP">
    <atom iupac_name="C" cns_name="C" dyana_name="C"/>
    <atom iupac_name="CA" cns_name="CA" dyana_name="CA"/>
    <atom iupac_name="CB" cns_name="CB" dyana_name="CB"/>
    <atom iupac_name="H" cns_name="HN" dyana_name="HN"/>
    <atom iupac_name="H1" cns_name="HT3" dyana_name="H1"/>
    <atom iupac_name="H2" cns_name="HT2" dyana_name="H2"/>
    <atom iupac_name="H3" cns_name="HT1" dyana_name="H3"/>
    <atom iupac_name="HA" cns_name="HA" dyana_name="HA"/>
    <atom iupac_name="HB2" cns_name="HB2" dyana_name="HB2"/>
    <atom iupac_name="HB3" cns_name="HB1" dyana_name="HB3"/>
    <atom iupac_name="SG" cns_name="SG" dyana_name="SG"/>
    <atom iupac_name="CAO" cns_name="CAO" dyana_name="CAO"/>
    <atom iupac_name="OAN" cns_name="OAN" dyana_name="OAN"/>
    <atom iupac_name="CAD" cns_name="CAD" dyana_name="CAD"/>
    <atom iupac_name="HAD" cns_name="HAD" dyana_name="HAD"/>
    <atom iupac_name="CAC" cns_name="CAC" dyana_name="CAC"/>
    <atom iupac_name="HAC" cns_name="HAC" dyana_name="HAC"/>
    <atom iupac_name="CAB" cns_name="CAB" dyana_name="CAB"/>
    <atom iupac_name="HAB1" cns_name="HAB1" dyana_name="HAB1"/>
    <atom iupac_name="HAB2" cns_name="HAB2" dyana_name="HAB2"/>
    <atom iupac_name="HAB3" cns_name="HAB3" dyana_name="HAB3"/>
    <atom iupac_name="OAA" cns_name="OAA" dyana_name="OAA"/>
    <atom iupac_name="HAA" cns_name="HAA" dyana_name="HAA"/>
    <atom iupac_name="CAE" cns_name="CAE" dyana_name="CAE"/>
    <atom iupac_name="HAE" cns_name="HAE" dyana_name="HAE"/>
    <atom iupac_name="CAF" cns_name="CAF" dyana_name="CAF"/>
    <atom iupac_name="HAF1" cns_name="HAF1" dyana_name="HAF1"/>
    <atom iupac_name="HAF2" cns_name="HAF2" dyana_name="HAF2"/>
    <atom iupac_name="NAP" cns_name="NAP" dyana_name="NAP"/>
    <atom iupac_name="HAP" cns_name="HAP" dyana_name="HAP"/>
    <atom iupac_name="CAQ" cns_name="CAQ" dyana_name="CAQ"/>
    <atom iupac_name="CAS" cns_name="CAS" dyana_name="CAS"/>
    <atom iupac_name="OAX" cns_name="OAX" dyana_name="OAX"/>
    <atom iupac_name="OAT" cns_name="OAT" dyana_name="OAT"/>
    <atom iupac_name="HAT" cns_name="HAT" dyana_name="HAT"/>
    <atom iupac_name="CAG" cns_name="CAG" dyana_name="CAG"/>
    <atom iupac_name="SAH" cns_name="SAH" dyana_name="SAH"/>
    <atom iupac_name="CAI" cns_name="CAI" dyana_name="CAI"/>
    <atom iupac_name="HAI1" cns_name="HAI1" dyana_name="HAI1"/>
    <atom iupac_name="HAI2" cns_name="HAI2" dyana_name="HAI2"/>
    <atom iupac_name="CAJ" cns_name="CAJ" dyana_name="CAJ"/>
    <atom iupac_name="HAJ1" cns_name="HAJ1" dyana_name="HAJ1"/>
    <atom iupac_name="HAJ2" cns_name="HAJ2" dyana_name="HAJ2"/>
    <atom iupac_name="NAK" cns_name="NAK" dyana_name="NAK"/>
    <atom iupac_name="HAK" cns_name="HAK" dyana_name="HAK"/>
    <atom iupac_name="CAL" cns_name="CAL" dyana_name="CAL"/>
    <atom iupac_name="HAL" cns_name="HAL" dyana_name="HAL"/>
    <atom iupac_name="NAM" cns_name="NAM" dyana_name="NAM"/>
    <atom iupac_name="HAM" cns_name="HAM" dyana_name="HAM"/>
    <atom iupac_name="HT" cns_name="None" dyana_name="None"/>
    <atom iupac_name="N" cns_name="N" dyana_name="N"/>
    <atom iupac_name="NT" cns_name="None" dyana_name="None"/>
    <atom iupac_name="O" cns_name="O" dyana_name="O"/>
    <atom iupac_name="O'" cns_name="OT1" dyana_name="O"/>
    <atom iupac_name="O''" cns_name="OT2" dyana_name="OXT"/>
  </residue>
</conversion_table>
```



Annexe 2.6: topallhdg5.3.pro

```
(...)

! IMP - mass for new atom types
MASS CHAE  12.0110
MASS CHAF  12.0110
MASS NHAP  14.0067
MASS HHAP   1.0080
MASS CHAQ  12.0110
MASS CHAS  12.0110
MASS CHAG  12.0110
MASS OHAX  15.9994
MASS OHAT  15.9994
MASS CHAL  12.0110
MASS NHAM  14.0067
MASS HHAM   1.0080

Residue IMP
GROUP
 ATOM O    TYPE=O    CHARGE=-0.457 END
 ATOM C    TYPE=C    CHARGE= 0.208 END
 ATOM CA   TYPE=CH1E CHARGE= 0.100 END
 ATOM HA   TYPE=HA   CHARGE= 0.013 END
 ATOM N    TYPE=NH1  CHARGE= 0.359 END
 ATOM HN   TYPE=H    CHARGE= 0.016 END
 ATOM CB   TYPE=CH2E CHARGE= 0.086 END
 ATOM HB1  TYPE=HA   CHARGE= 0.013 END
 ATOM HB2  TYPE=HA   CHARGE= 0.013 END
 ATOM SG   TYPE=SM   CHARGE= 0.051 END
 ATOM CAO  TYPE=C    CHARGE= 0.208 END
 ATOM OAN  TYPE=O    CHARGE=-0.457 END
 ATOM CAD  TYPE=CH1E CHARGE= 0.100 END
 ATOM HAD  TYPE=HA   CHARGE= 0.013 END
 ATOM CAC  TYPE=CH1E CHARGE= 0.101 END
 ATOM HAC  TYPE=HA   CHARGE= 0.013 END
 ATOM CAB  TYPE=CH3E CHARGE= 0.085 END
 ATOM HAB1 TYPE=HA   CHARGE= 0.013 END
 ATOM HAB2 TYPE=HA   CHARGE= 0.013 END
 ATOM HAB3 TYPE=HA   CHARGE= 0.013 END
 ATOM OAA  TYPE=OH1  CHARGE=-0.109 END
 ATOM HAA  TYPE=H    CHARGE= 0.040 END
 ATOM CAE  TYPE=CHAE CHARGE= 0.101 END !new type
 ATOM HAE  TYPE=HA   CHARGE= 0.013 END
 ATOM CAF  TYPE=CHAF CHARGE= 0.085 END !new type
 ATOM HAF1 TYPE=HA   CHARGE= 0.013 END
 ATOM HAF2 TYPE=HA   CHARGE= 0.013 END
 ATOM NAP  TYPE=NHAP CHARGE=-0.165 END !new type
 ATOM HAP  TYPE=HHAP CHARGE= 0.016 END !new type
 ATOM CAQ  TYPE=CHAQ CHARGE= 0.086 END !new type
 ATOM CAS  TYPE=CHAS CHARGE= 0.208 END !new type
 ATOM OAX  TYPE=OHAX CHARGE=-0.457 END !new type
 ATOM OAT  TYPE=OHAT CHARGE=-0.110 END !new type
 ATOM HAT  TYPE=H    CHARGE= 0.040 END
 ATOM CAG  TYPE=CHAG CHARGE= 0.005 END !new type
 ATOM SAH  TYPE=SM   CHARGE= 0.051 END
 ATOM CAI  TYPE=CH2E CHARGE= 0.086 END
 ATOM HAI1 TYPE=HA   CHARGE= 0.013 END
 ATOM HAI2 TYPE=HA   CHARGE= 0.013 END
 ATOM CAJ  TYPE=CH2E CHARGE= 0.086 END
 ATOM HAJ1 TYPE=HA   CHARGE= 0.013 END
 ATOM HAJ2 TYPE=HA   CHARGE= 0.013 END
 ATOM NAK  TYPE=NH1  CHARGE=-0.165 END
 ATOM HAK  TYPE=H    CHARGE= 0.016 END
 ATOM CAL  TYPE=CHAL CHARGE= 0.030 END !new type
 ATOM HAL  TYPE=HA   CHARGE= 0.013 END
 ATOM NAM  TYPE=NHAM CHARGE=-0.513 END !new type
 ATOM HAM  TYPE=HHAM CHARGE= 0.016 END !new type

 BOND N    HN
 BOND CA   N     BOND CA   HA
 BOND CA   CB    BOND CB   HB1   BOND CB   HB2
 BOND CB   SG
 BOND CAO  SG    BOND CAO  OAN
 BOND CAD  CAO   BOND CAD  HAD   BOND CAD  CAC
 BOND CAE  CAD
 BOND CAC  HAC   BOND CAC  CAB   BOND CAC  OAA
 BOND CAB  HAB1  BOND CAB  HAB2  BOND CAB  HAB3
 BOND OAA  HAA
 BOND CAE  HAE   BOND CAE  CAF   BOND CAE  NAP
 BOND CAF  HAF1  BOND CAF  HAF2
 BOND CAG  CAF
 BOND NAP  HAP
 BOND CAQ  NAP   BOND CAQ  CAS   BOND CAQ  CAG
 BOND CAS  OAX   BOND CAS  OAT
 BOND OAT  HAT1
 BOND CAG  SAH
 BOND CAI  SAH   BOND CAI  HAI1  BOND CAI  HAI2  BOND CAI  CAJ
 BOND CAJ  HAJ1  BOND CAJ  HAJ2  BOND CAJ  NAK
 BOND NAK  HAK   BOND NAK  CAL
 BOND CAL  HAL   BOND CAL  NAM
 BOND NAM  HAM
 BOND CA   C
 BOND C    O

 IMPROPER HA  N  C    CB                !chirality CA
 IMPROPER HB1 HB2 CA SG                 !stereo CB
 IMPROPER CAO  SG   OAN  CAD            !CAO
 IMPROPER HAD     CAO  CAC  CAE         !CAD
 IMPROPER HAC  CAD  OAA  CAB            !CAC
 IMPROPER HAB1 HAB2 CAC HAB3            !methyl CAB
 IMPROPER HAF1 HAF2 CAE CAG             !stereo CAF
 IMPROPER HAE  NAP  CAD CAF             !chirality CAE
 IMPROPER NAP  CAE  HAP  CAQ
 IMPROPER CAQ  NAP  CAS  CAG
 IMPROPER CAS  OAT  OAX  CAQ
 IMPROPER CAG  CAF  SAH  CAQ
 IMPROPER  HAI1 HAI2 SAH  CAJ           !stereo CAI
 IMPROPER  HAJ1 HAJ2 CAI  NAK           !stereo CAJ
 IMPROPER NAK  CAJ  HAK  CAL
 IMPROPER CAL  NAK  HAL  NAM
 IMPROPER HAM  NAM  CAL  HAL

 DIHEDRAL CB   CA  C    O
 DIHEDRAL C    CA  N    HN
 DIHEDRAL SG   CB  CA   C
 DIHEDRAL CA   CB  SG   CAO
 DIHEDRAL CAD  CAO  SG   CB
 DIHEDRAL CAE  CAD  CAO  SG
 DIHEDRAL OAA  CAC  CAD  CAO
 DIHEDRAL NAP  CAE  CAD  CAO
 DIHEDRAL HAB3 CAB  CAC  CAD
 DIHEDRAL CAD  CAC  OAA  HAA
 DIHEDRAL CAG  CAF  CAE  CAD
 DIHEDRAL CAD  CAE  NAP  CAQ
 DIHEDRAL CAE  CAF  CAG  SAH
 DIHEDRAL CAG  CAQ  NAP  CAE
 DIHEDRAL OAT  CAS  CAQ  NAP
 DIHEDRAL SAH  CAG  CAQ  NAP
 DIHEDRAL CAQ  CAS  OAT  HAT1
 DIHEDRAL CAF  CAG  SAH  CAI
 DIHEDRAL CAJ  CAI  SAH  CAG
 DIHEDRAL NAK  CAJ  CAI  SAH
 DIHEDRAL CAI  CAJ  NAK  CAL
 DIHEDRAL NAM  CAL  NAK  CAJ

 DONO HN   N
 ACCE O    C
end
```

Annexe 2.7: parallhdg5.3.pro

```
(...)
BOND C      SM     1000.000 {sd=   0.001}   1.750      ! for IMP
BOND CHAE   CH1E   1000.000 {sd=   0.001}   1.530      ! for IMP
BOND CHAE   HA     1000.000 {sd=   0.001}   1.090      ! for IMP
BOND CHAE   CHAF   1000.000 {sd=   0.001}   1.520      ! for IMP
BOND CHAE   NHAP   1000.000 {sd=   0.001}   1.470      ! for IMP
BOND CHAF   HA     1000.000 {sd=   0.001}   1.090      ! for IMP
BOND CHAG   CHAF   1000.000 {sd=   0.001}   1.390      ! for IMP
BOND NHAP   HHAP   1000.000 {sd=   0.001}   1.000      ! for IMP
BOND CHAQ   NHAP   1000.000 {sd=   0.001}   1.330      ! for IMP
BOND CHAQ   CHAS   1000.000 {sd=   0.001}   1.330      ! for IMP
BOND CHAQ   CHAG   1000.000 {sd=   0.001}   1.330      ! for IMP
BOND CHAS   OHAX   1000.000 {sd=   0.001}   1.230      ! for IMP
BOND CHAS   OHAT   1000.000 {sd=   0.001}   1.360      ! for IMP
BOND OHAT   H      1000.000 {sd=   0.001}   1.000      ! for IMP
BOND CHAG   SM     1000.000 {sd=   0.001}   1.750      ! for IMP
BOND NH1    CHAL   1000.000 {sd=   0.001}   1.330      ! for IMP
BOND CHAL   HA     1000.000 {sd=   0.001}   1.090      ! for IMP
BOND CHAL   NHAM   1000.000 {sd=   0.001}   1.340      ! for IMP
BOND NHAM   HHAM   1000.000 {sd=   0.001}   1.000      ! for IMP

(...)
ANGLe CH1E  CH2E   SM      500.000 {sd=   0.031}   113.000    ! for IMP
ANGLe CH2E  SM     C       500.000 {sd=   0.031}   100.000    ! for IMP
ANGLe SM    C      O       500.000 {sd=   0.031}   124.000    ! for IMP
ANGLe SM    C      CH1E    500.000 {sd=   0.031}   115.000    ! for IMP
ANGLe C     CH1E   CHAE    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe HA    CH1E   CHAE    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CH1E  CH1E   CHAE    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CH1E  CHAE   HA      500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CH1E  CHAE   CHAF    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CH1E  CHAE   NHAP    500.000 {sd=   0.031}   111.000    ! for IMP
ANGLe HA    CHAE   CHAF    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe HA    CHAE   NHAP    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CHAF  CHAE   NHAP    500.000 {sd=   0.031}   104.000    ! for IMP
ANGLe CHAE  CHAF   HA      500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CHAE  CHAF   CHAG    500.000 {sd=   0.031}   104.000    ! for IMP
ANGLe HA    CHAF   HA      500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe HA    CHAF   CHAG    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CHAE  NHAP   HHAP    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CHAE  NHAP   CHAQ    500.000 {sd=   0.031}   104.000    ! for IMP
ANGLe HHAP  NHAP   CHAQ    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe NHAP  CHAQ   CHAS    500.000 {sd=   0.031}   126.000    ! for IMP
ANGLe NHAP  CHAQ   CHAG    500.000 {sd=   0.031}   115.000    ! for IMP
ANGLe CHAS  CHAQ   CHAG    500.000 {sd=   0.031}   120.000    ! for IMP
ANGLe CHAQ         CHAS    OHAX    500.000 {sd=   0.031}   121.000    ! for IMP
ANGLe CHAQ         CHAS    OHAT    500.000 {sd=   0.031}   115.000    ! for IMP
ANGLe OHAX         CHAS    OHAT    500.000 {sd=   0.031}   124.000    ! for IMP
ANGLe CHAS  OHAT   H       500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe CHAF  CHAG   CHAQ    500.000 {sd=   0.031}   108.000    ! for IMP
ANGLe CHAF  CHAG   SM      500.000 {sd=   0.031}   120.000    ! for IMP
ANGLe CHAQ         CHAG    SM      500.000 {sd=   0.031}   120.000    ! for IMP
ANGLe CHAG  SM     CH2E    500.000 {sd=   0.031}   100.000    ! for IMP
ANGLe CH2E  NH1    CHAL    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe H     NH1    CHAL    500.000 {sd=   0.031}   109.500    ! for IMP
ANGLe NH1   CHAL   HA      500.000 {sd=   0.031}   115.000    ! for IMP
ANGLe NH1   CHAL   NHAM    500.000 {sd=   0.031}   115.000    ! for IMP
ANGLe HA    CHAL   NHAM    500.000 {sd=   0.031}   115.000    ! for IMP
ANGLe CHAL  NHAM   HHAM    500.000 {sd=   0.031}   123.000    ! for IMP

(...)
IMPRoper    C      SM      O       CH1E    500.000 {sd=   0.031}0      0.0000 ! for IMP
IMPRoper    NHAP   CHAE    HHAP    CHAQ    500.000 {sd=   0.031}0      39.8000      ! for IMP
IMPRoper    CHAQ   NHAP    CHAS    CHAG    500.000 {sd=   0.031}0      0.0000 ! for IMP
IMPRoper    CHAS   OHAT    OHAX    CHAQ    500.000 {sd=   0.031}0      0.0000 ! for IMP
```

Annexe 2.7 - continued

```
IMPRoper     CHAG  CHAF  SM     CHAQ  500.000 {sd=    0.031}0        0.0000          ! for IMP
IMPRoper     CH2E  SM    HA     HA      500.000 {sd=    0.031}0        35.2640         ! for IMP
IMPRoper     NH1   CH2E  H      CHAL   500.000 {sd=    0.031}0        35.2640         ! for IMP
IMPRoper     CHAL  NH1   HA     NHAM 500.000 {sd=    0.031}0        0.0150          ! for IMP
IMPRoper     HA  HA  CH1E SM     500.00 {sd=    0.031}          0   -72.0234   ! for IMP
IMPRoper     HA  C   CH1E CHAE   500.00 {sd=    0.031}          0   66.1640    ! for IMP
IMPRoper     HA  OH1 CH1E CHAE   500.00 {sd=    0.031}          0   66.1640    ! for IMP
IMPRoper     CHAG  CHAE HA  HA     500.00 {sd=    0.031}          0   -70.4072   ! for IMP
IMPRoper     HA  NHAP CH1E  CHAF   500.00 {sd=    0.031}          0   66.1640    ! for IMP
IMPRoper     HHAM NHAM CHAL  HA     500.00 {sd=    0.031}     0   0.0000     ! for IMP

! CR3-CR3 sidechain dihedrals
(...)
DIHEdral  C           CH1E  NH1   H       2.00   3  0.0000 ! for IMP
DIHEdral  SM          CH2E  CH1E  C       2.00   3  0.0000 ! for IMP
DIHEdral  CH1E        CH2E  SM    C       2.00   3  0.0000 ! for IMP
DIHEdral  CH1E        C     SM    CH2E    2.00   3  0.0000 ! for IMP
DIHEdral  CHAE        CH1E  C     SM      2.00   6  0.0000 ! for IMP
DIHEdral  OH1         CH1E  CH1E  C       2.00   3  0.0000 ! for IMP
DIHEdral  NHAP        CHAE  CH1E  C       2.00   3  0.0000 ! for IMP
DIHEdral  HA          CH3E  CH1E  CH1E    2.00   3  0.0000 ! for IMP
DIHEdral  CH1E        CH1E  OH1   H       2.00   3  0.0000 ! for IMP
DIHEdral  CHAG        CHAF  CHAE  CH1E    2.00   3  0.0000 ! for IMP
DIHEdral  CH1E        CHAE  NHAP  CHAQ    2.00   3  0.0000 ! for IMP
DIHEdral  CHAE        CHAF  CHAG  SM      2.00   6  0.0000 ! for IMP
DIHEdral  CHAG        CHAQ  NHAP  CHAE    2.00   3  0.0000 ! for IMP
DIHEdral  OHAT        CHAS  CHAQ  NHAP    2.00   2  180.0000 ! for IMP
DIHEdral  SM          CHAG  CHAQ  NHAP    2.00   2  180.0000 ! for IMP
DIHEdral  CHAQ        CHAS  OHAT  H       2.00   2  180.0000 ! for IMP
DIHEdral  CHAF        CHAG  SM    CH2E    2.00   3  0.0000 ! for IMP
DIHEdral  CH2E        CH2E  SM    CHAG    2.00   3  0.0000 ! for IMP
DIHEdral  NH1         CH2E  CH2E  SM      2.00   3  0.0000 ! for IMP
DIHEdral  CH2E        CH2E  NH1   CHAL    2.00   3  0.0000 ! for IMP
DIHEdral  NHAM        CHAL  NH1   CH2E    2.00   3  0.0000 ! for IMP


{* nonbonding parameter section *}
(...)
NONBonded  CHAE   $VE $VR_C   $VE $VR14_C       ! for IMP
NONBonded  CHAF   $VE $VR_C   $VE $VR14_C       ! for IMP
NONBonded  CHAQ   $VE $VR_C   $VE $VR14_C       ! for IMP
NONBonded  CHAS   $VE $VR_C   $VE $VR14_C       ! for IMP
NONBonded  CHAG   $VE $VR_C   $VE $VR14_C       ! for IMP
NONBonded  CHAL   $VE $VR_C   $VE $VR14_C       ! for IMP
NONBonded  NHAP   $VE $VR_N   $VE $VR14_N       ! for IMP
NONBonded  NHAM   $VE $VR_N   $VE $VR14_N       ! for IMP
NONBonded  HHAP   $VE $VR_HH  $VE $VR14_HH      ! for IMP
NONBonded  HHAM   $VE $VR_HH  $VE $VR14_HH      ! for IMP
NONBonded  OHAX   $VE $VR_O   $VE $VR14_O       ! for IMP
NONBonded  OHAT   $VE $VR_O   $VE $VR14_O       ! for IMP

NONBONDED CHAE  0.10000 3.29633 0.10000 3.02906    ! for IMP
NONBONDED CHAF  0.10000 3.29633 0.10000 3.02906    ! for IMP
NONBONDED NHAP  0.10000 2.67270 0.10000 2.40543    ! for IMP
NONBONDED HHAP  0.10000 1.42544 0.10000 1.15817    ! for IMP
NONBONDED CHAQ  0.10000 3.29633 0.10000 3.02906    ! for IMP
NONBONDED CHAS  0.10000 3.29633 0.10000 3.02906    ! for IMP
NONBONDED OHAX  0.10000 2.58361 0.10000 2.31634    ! for IMP
NONBONDED OHAT  0.10000 2.58361 0.10000 2.31634    ! for IMP
NONBONDED CHAG  0.10000 3.29633 0.10000 3.02906    ! for IMP
NONBONDED CHAL  0.10000 3.29633 0.10000 3.02906    ! for IMP
NONBONDED NHAM  0.10000 2.67270 0.10000 2.40543    ! for IMP
NONBONDED HHAM  0.10000 1.42544 0.10000 1.15817    ! for IMP
(...)
```

RÉSUMÉ

L'étape finale de biosynthèse du peptidoglycane est catalysée par les D,D-transpeptidases (PBPs), l'une des cibles principales des antibiotiques de type β-lactame. Récemment, il a été montré qu'une nouvelle classe d'enzymes, les L,D-transpeptidases (LDts), permet de contourner l'inhibition des PBPs. Ces LDts ont été identifiées tant dans des bactéries résistantes aux β-lactames que dans des formes dormantes de *Mycobacterium tuberculosis*. Les seuls β-lactames capables de les inhiber, les carbapénèmes, forment une liaison covalente avec la cystéine catalytique des LDts. Ni le mécanisme de cette inactivation, ni la spécificité de ces enzymes pour les carbapénèmes ne sont toutefois expliqués à ce jour. Le but du présent travail consiste en l'investigation par RMN du mécanisme d'acylation des LDts par ces antibiotiques. Dans ce contexte, la première partie de cette thèse s'intéresse à la compréhension actuelle de l'émergence de ce phénomène de résistance. La seconde partie traite des principes de la RMN et des implémentations développées pour étudier la structure, la thermodynamique et la dynamique des LDts. La troisième et dernière partie démontre le succès de l'approche RMN dans l'étude des diverses étapes de la réaction d'acylation, à travers une étude détaillée de l'apoenzyme, de complexes non covalents avec différents β-lactames, et de l'enzyme acylée par un carbapénème. Au cours de cette étude, la structure du site actif de l'apoenzyme de *Bacillus subtilis* a été affinée par rapport à une étude cristallographique antérieure. Pour cette enzyme et son pendant chez *Enterococcus faecium*, nous avons démontré que la spécificité pour les carbapénèmes n'intervient pas au stade de la formation du complexe non covalent. Pour finir, la formation de la liaison covalente entre LDt et carbapénème induit un réarrangement conformationnel substantiel et une augmentation de la flexibilité de l'enzyme.

ABSTRACT

The final cross-linking step of the peptidoglycan synthesis is usually catalyzed by D,D-transpeptidases (PBPs), one of the main targets of β-lactam antibiotics. Recently, it was shown that these PBPs can be by-passed by a novel class of enzymes, the L,D-transpeptidases (LDts), identified in β-lactam-resistant bacteria as well as in dormant forms of *Mycobacterium tuberculosis*. The only β-lactams enable to inactivate these enzymes belong to the carbapenem class. The β-lactam ring of this antibiotic family then covalently binds the catalytic cysteine of the LDt. Neither the mechanism of this reaction nor the specificity for carbapenems are yet understood. The aim of the present work is to investigate the acylation mechanism of LDts with carbapenems by NMR. In this context, the first part of this thesis focuses on the current biological understanding of the emergence of this resistance pathway. The second part deals with the NMR principles and the implementations developed to study the structure, thermodynamics and dynamics of LDts. The third part demonstrates that NMR is successful in studying all the steps of the acylation reaction. For this purpose, the LDt apoenzyme, the non-covalent complex with various β-lactams, and the LDt-carbapenem acylenzyme were thoroughly investigated. The structure of the active site of the *Bacillus subtilis* apoenzyme was refined with respect to a previous crystallographic study. For the latter and the *Enterococcus faecium* enzymes, we showed that the carbapenem specificity does not occur at the stage of the non-covalent binding. In contrast to non-covalent interactions, the formation of the covalent bond between LDts and carbapenems induces substantial conformational rearrangement and increased flexibility in the enzyme.

Key words: peptidoglycan, L,D-transpeptidation, NMR, β-lactam antibiotic, carbapenem